nature

THE WAY AHEAD

Lessons from the International Polar Year

NANOSCALE FRICTION
Laying down the law

ALZHEIMER'S DISEASE
Prion protein implicated

NEUROETHICS
When brain meets machine

The way ahead for polar science

The fourth International Polar Year is coming to an end. But the Arctic and Antarctic must remain top scientific priorities.

Next month's completion of the International Polar Year (IPY) is an occasion worth celebrating. Over two full Arctic and Antarctic field seasons, from March 2007 to March 2009, this global meta-project has successfully coordinated the efforts of some 50,000 scientists — including many from poorer nations, and from countries with a relatively new interest in polar research. It has also greatly advanced our still-sketchy understanding of physical, chemical and biological processes near the poles (see page 1072).

But the IPY also dramatized the polar regions' rapidly increasing connections to the rest of the globe. At the time of the last such exercise, the International Geophysical Year of 1957–58, the poles still seemed almost as remote as Mars. Today, they are feeling the effects of global climate change more intensely than anywhere else on Earth — and are fast becoming a new frontier for economic and political rivalries. Scientists and policy-makers should therefore think of this IPY not as a glorious endeavour that is now over and done with, but as a foundation on which to keep building.

The first priority is to ensure that the achievements of the IPY are exploited to the fullest. For example, the IPY programme has still not created a freely accessible archive for all the data and scientific findings harvested under its banner — information that will be in high demand from physical oceanographers, marine biologists, climate scientists and many others. Such an archive should be created as soon as possible, with funding coming from all the IPY nations relative to their economic strength. Nor has there yet been any systematic assessment of the IPY's scientific findings in a form that would be meaningful to decision-makers and the public at large. The International Arctic Science Committee and the Scientific Committee on Antarctic Research, the main organizations for polar science, are best placed to organize such an assessment. And Norway, which is a member of almost all the polar governance treaties and organizations, has promised to take the political lead in initiating the effort. It deserves the support of all the nations and organizations that participated in the IPY.

Looking to the future, meanwhile, polar stakeholders should vigorously pursue plans to create permanent observation networks at the poles. As the IPY itself underscored, for example, the rapidly warming climate creates an urgent need for ongoing ground observations of ice thickness and ice movement, as well as a new satellite infrastructure to monitor the cryosphere from space. The global economic downturn will obviously make it difficult for governments to fund such networks. But it is worth noting that at least some related projects will qualify for stimulus money — one of them being the Alaska Region Research Vessel, the new US flagship for polar research.

In parallel, the various polar stakeholders should pursue a sustainable, ecosystem-based management regime for both polar regions. New or improved international regulatory arrangements are needed in many areas, from fishing and shipping to pollution control and climate mitigation. And especially in the Arctic, which does not enjoy the international protections afforded the southern continent under the 50-year-old Antarctic Treaty, the rule-making process needs to encompass a hard, systematic look at economic and geopolitical governance. National directives on the Arctic issued by Russia, the United States, the European Union, Canada, Denmark and Norway suggest that future Arctic governance will be a delicate matter indeed. Add in the interest being shown by new players such as China and Brazil, and the need for concerted action is clear.

> "The first priority is to ensure that the achievements of the IPY are exploited to the fullest."

Finally, the new economic and political interest in the Arctic makes it all the more important that scientists keep alive the IPY's spirit of transnational cooperation. The poles have always drawn people with grand ambitions — not just scientists, but adventurers, military operatives and entrepreneurs of all stripes. And they are fast becoming more attractive, with all the potential for rivalry and discord that implies. Scientific collaborations will not stop that process. But they can keep alive important avenues of international communication, as they have in the past. And they can help nations remember that what happens at the poles is inseparable from what happens in the rest of the world. ■

Europe's GM quandary

A political impasse over transgenic crops has left the European Commission with no good options.

It seems like a lose–lose situation for Europe's environment commissioner, Stavros Dimas, whose thankless task it is to enforce the legislation that governs genetically modified (GM) crops. Next week, the council of environment ministers from the 27 member states of the European Union (EU) will vote on whether the insect-resistant maize MON801, the only GM crop approved for cultivation in Europe, should still be allowed. But a firm decision in favour or against will require what is known as a qualified majority, representing at least 62% of the EU population, and this degree of consensus seems unlikely given Europe's deep divisions over GM crops (see *Nature* **457,** 946; 2009). So the decision is likely to get bumped up to the European Commission itself — which will be bound by its own rules to decide in favour, thereby unleashing political fireworks.

That the council should be involved at all in such a technical matter speaks volumes for how contentious this issue is. In 2004,

after six years of squabbling, the EU member states formally implemented a directive that allows the cultivation of GM crops — albeit with the toughest environmental and health safeguards in the world. The directive, in turn, is implemented through the European Food Safety Authority, which relies on scientific experts to assess whether a crop should be approved for cultivation.

In most other scientific or technical matters — approving a new medical device, say, or authorizing a new toxicity test — the experts' assessment is rubber-stamped by the appropriate regulatory committee. In effect, science has the final say. But that is not the case for GM crops. Four EU member states, Austria, Hungary, Greece and France, have now compiled new scientific evidence showing — or so they claim — that MON801 can endanger the environment. The European Food Safety Authority disagrees. And the EU regulatory committee is deadlocked on what to do about it.

Thus the involvement of the environment ministers. However, insiders are anticipating that they, too, will be politically deadlocked — even though, according to the directive, the scientific advice should determine the outcome. The decision will therefore have to be made by the commission, which is obliged to follow the scientists' advice and vote in favour of continued cultivation. But, being composed of unelected officials, it will undoubtedly be accused of anti-democratic action if it does.

The whole problem might be solved if countries opposed to GM crops could simply opt out of the legislation. But that would violate a core philosophy of the EU, which is the free movement of goods and people between all countries. Tinkering with the existing law is no solution, either: GM crops currently have too little support in Europe for any form of legislation to be robust.

So the only other option is to wait: let the current stalemate continue until the public opposition to GM crops begins to fade. In some European countries, GM crops have brought agricultural benefits and public opposition is relatively light. And surveys suggest that the European public is slowly starting to accept the idea. GM crops, as far as science can tell, are not harmful, and if, as is to be expected, Europe's consumers can benefit from cheaper, better food, or can be convinced of broader benefits amid a global food crisis, then opposition will decline. Ultimately, the onus is on manufacturers to deliver the products that will help to shift that political deadlock. ■

It's good to blog

More researchers should engage with the blogosphere, including authors of papers in press.

Is blogging a part of science, journalism or public discourse? In fact it may be all of these — an ambiguity that can sometimes leave scientists feeling uncertain about the rules of the game.

Imagine, for example, a case in which *Nature*'s blog The Great Beyond highlights new scientific results presented at a conference on climate. That blog entry then stimulates an online debate, with climate sceptics interpreting the results their way, and others firing off rebuttals. Imagine also that the work is described in a paper that had been accepted, but not published, by *Nature*. The authors of the paper want to enter the fray, but feel inhibited from doing so because of the embargo imposed by *Nature* and many other journals on communication by authors to the media ahead of publication. And why was *Nature* blogging their work anyway, ahead of its publication?

This scenario highlights a need for clarification about *Nature* publications' procedures, and about how embargoes apply to blogs. It also highlights more generally the potential importance of scientists engaging in the blogosphere.

All *Nature* journals maintain confidentiality about submitted papers, so that only the editors directly responsible for those papers know about them. Other staff — including the various publications' journalists — are usually informed about a paper only once it has been accepted, and with the proviso that they do not disseminate any information about it to external contacts or readers. Likewise, we ask that authors refrain from actively promoting their work to the media and public ahead of its publication. This embargo policy rests on the principle that scientists' and the public's best interests are served by press coverage of work that has been peer reviewed, and is available for others to see for themselves.

At the same time, however, our cardinal rule has always been to promote scientific communication. We have therefore never sought to prevent scientists from presenting their work at conferences, or from depositing first drafts of submitted papers on preprint servers. So if *Nature* journalists or those from any other publication should hear results presented at a meeting, or find them on a preprint server, the findings are fair game for coverage — even if that coverage is ahead of the paper's publication. This is not considered a breaking of *Nature*'s embargo. Nor is it a violation if scientists respond to journalists' queries in ensuring that the facts are correct — so long as they don't actively promote media coverage.

> **"There are societal debates that have much to gain from the uncensored voices of researchers."**

The blogosphere differs from mass media and specialized media in many respects, but the same considerations apply in disseminating new scientific results there. Authors of papers in press have the right to correct misrepresentations and to point to results that will appear in a paper. But a full discussion should await the paper's publication.

Indeed, researchers would do well to blog more than they do. The experience of journals such as *Cell* and *PLoS ONE*, which allow people to comment on papers online, suggests that researchers are very reluctant to engage in such forums. But the blogosphere tends to be less inhibited, and technical discussions there seem likely to increase.

Moreover, there are societal debates that have much to gain from the uncensored voices of researchers. A good blogging website consumes much of the spare time of the one or several fully committed scientists that write and moderate it. But it can make a difference to the quality and integrity of public discussion. ■

Discuss this Editorial online at http://tinyurl.com/c6zoq6.

RESEARCH HIGHLIGHTS

Prion hijackers

Nature Cell Biol. doi:10.1038/ncb1841 (2009)

Prions that cause brain disease could spread between cells by hijacking a system cells may be using as a way of communicating, Chiara Zurzolo of the Pasteur Institute in Paris and her colleagues report. This may be how infectious prions infiltrate the central nervous system.

The researchers found that prions labelled with a fluorescent protein shuttled from one nerve cell to another by travelling inside nanotubes that connect these cells.

Treatment with a chemical that prevents nanotube formation halted the transfer of prions. Prions also moved from immune-system cells called dendritic cells to neurons.
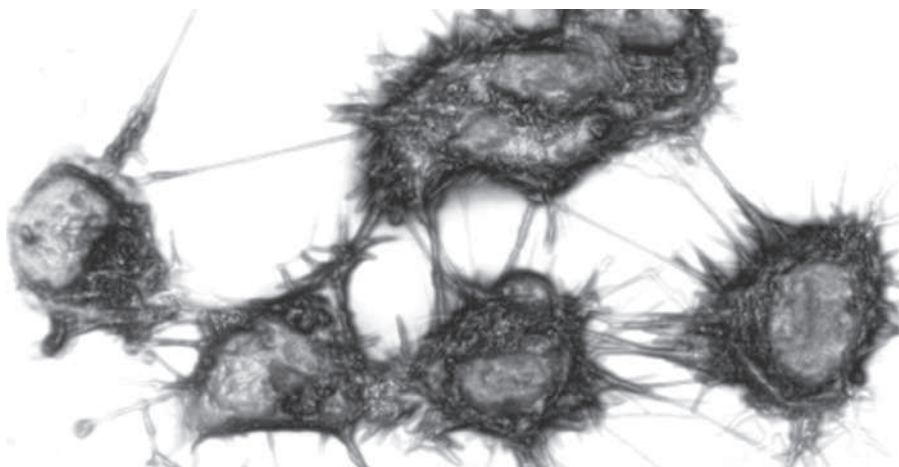
K. GOUSSET ET AL.

PLANT SCIENCES
Rust resistant

Science doi:10.1126/science.1166453 (2009)

Why some rust-resistant plant genes maintain their potency as the fungus co-evolves has long baffled crop researchers. For instance, over the past 50 years leaf rust and stripe rust have become no more virulent against wheat plants carrying the gene *Lr34*. But other resistance genes are typically rendered useless by these pathogens within three to five years.

Lagudah Evans at CSIRO Plant Industry in Canberra, Beat Keller of the University of Zurich in Switzerland and their colleagues recently studied *Lr34*'s sequence. The team discovered that the crucial DNA sequence encodes a type of membrane protein called an ATP-binding cassette, or 'ABC' transporter.

Scientists who come across new resistance genes will be able to screen them for similar sequences and, on that basis, predict whether the resistance the new gene confers is likely to last, the authors say.

MOLECULAR BIOLOGY
Adaptation

Nature Biotechnol. doi:10.1038/nbt.1525 (2009)

A new way to silence gene expression has been devised. It differs from two common methods — antisense and RNA interference (RNAi) — in that it uses short synthetic nucleic-acid molecules known as U1 Adaptors that work inside the nucleus.

These tether another molecule called U1 snRNP splicing factor to messenger RNA molecules that have just been made. The splicing factor inhibits the processing necessary for messenger RNAs to move to a ribosome, where the genetic information they carry would be translated into proteins.

Sam Gunderson of Rutgers University in Piscataway, New Jersey, and his colleagues created U1 Adaptors that halved gene expression when added to cells at sub-nanomolar concentrations — a level of activity on a par with RNAi. Combining U1 Adaptors and RNAi further reduced gene expression.

ANIMAL BEHAVIOUR
The carriers' code

J. Exp. Biol. **212,** 499–505 (2009)

Although leaf-cutter ants leaving their nest often encounter leaf-laden colleagues coming the other way, the ant traffic never becomes gridlocked. Using an experimental nest of *Atta colombica*, Audrey Dussutour of the Paul Sabatier University in Toulouse, France, and her colleagues have determined the rules of the insect road.

When the ants were forced to cross a narrow bridge, they tended to form clusters of inbound and outbound foragers. Inbound clusters were headed by leaf-carrying ants and were almost always given priority; outbound ants tended to cross when there were no inbound ants.

The few inbound ants without loads followed behind slower, load-bearing members of their cluster. Had they raced ahead, head-to-head encounters would have caused twice the delay imposed by tailgating behaviour.
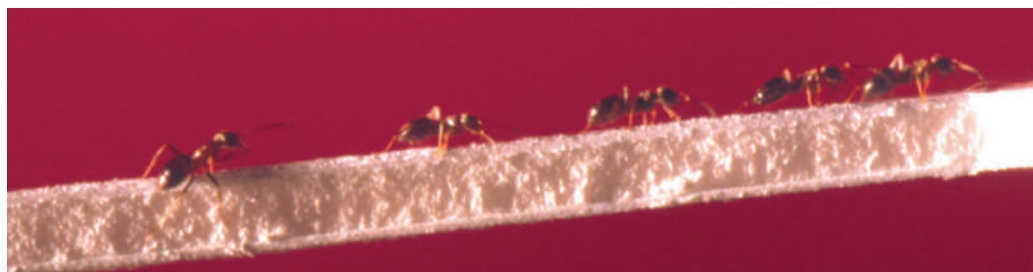
PHYSICS
Atomic quantum dots

Phys. Rev. Lett. **102,** 046805 (2009)

Assemblies of quantum dots — blobs of a few thousand atoms — with electrons shuttling between them could form miniature low-power computing circuits that do not require transistors. But these architectures are fragile and must operate at temperatures close to absolute zero to control the interactions of the electrons they confine.

A team led by Robert Wolkow at Canada's National Institute for Nanotechnology in Edmonton, Alberta, has discovered that single silicon atoms, sitting in an electron-doped silicon lattice that is blanketed with hydrogen, provide electronic structures with better properties than quantum dots.

The atoms can be joined in assemblies much like the dots. Besides being smaller, these are much more robust to external disturbance and can tunnel electrons controllably at room temperature.

OCEANOGRAPHY
Sky view

Geophys. Res. Lett. doi:10.1029/2008GL036422 (2009)

The localized sinking of large volumes of surface water to great depth has a crucial role in global ocean circulation and so in climate

CNRS PHOTOTHÈQUE

regulation. Buoys and ship-based sensors are normally used to measure the amount of water that sinks and how fast it does so, but generating such data from satellite readings would provide more complete coverage.

Marine Herrmann of the CNRS in Toulouse, France, and her colleagues have gone some way towards this by comparing satellite measurements of sea-surface elevation for the years 1994–2007 to a simulation of the Western Mediterranean Deep Water convection.

Sea level is always lower where convection takes place. Herrmann's study shows that it is sufficiently so off the southern coast of France (and by implication, many other sites around the world) for altimetry measurements to determine year-on-year convection changes.

NANOTECHNOLOGY
Etch a circuit sketch

Science **323,** 1026–1030 (2009)

The difficulty in designing nanoscale circuit boards lies in keeping electrons from leaving the conducting material through which they flow. With this in mind, Jeremy Levy of the University of Pittsburgh in Pennsylvania and his colleagues have devised 2-nanometre-wide circuits that confine electrons to the two dimensions of the chip's 'wiring' by trapping a gas of them at the interface of polar and nonpolar metal oxides.

Their circuits are made using a conducting atomic force microscope tip with positive voltage on the polar oxide, which changes the electronic properties of the oxide. Passing a tip with a negative voltage back over the circuits erases them. These processes can be repeated hundreds of times. The authors demonstrated the concept by building electronic components such as field-effect transistors.

MATERIALS SCIENCE
Better coats

Phys. Rev. Lett. **102,** 045003 (2009)

Sputtering, a common industrial method for coating surfaces, uses ions in a gas to knock metal atoms from a solid that then fly off to coat a target in a thin film. 'Self-sputtering' is a way to coat targets without the presence of a gas; at higher voltages some of the metal atoms are themselves ionized and return to their solid source, where they dislodge yet more metal atoms.

Now Joakim Andersson and André Anders of Lawrence Berkeley National Laboratory in California have created a sort of runaway self-sputtering. It uses pulses of extremely high voltage to make the ion current of the metal atoms exceed the applied electric current.

PALAEONTOLOGY
Flower power

J. Evol. Biol. **22,** 446–459 (2009)

Many palaeontologists have long thought that flowering plants and dinosaurs co-evolved because many species of both appeared during the Cretaceous period, 145 million to 65 million years ago. This now seems unlikely.

Richard Butler and his colleagues at London's Natural History Museum have mapped the species diversity of fossil finds encompassing 407 species of dinosaur (including those of diplodocus, pictured below) and more than 2,300 species of plant. They found no overall geographical correlation between the two data sets.

Instead, they learned that stegosaur diversity was negatively correlated with the diversity of flowering plants and positively correlated with that of non-flowering cycadophytes, which hints that the spiny-backed group ate cycadophytes.

NEUROSCIENCE
Mouse mapping

PLoS Biol. **7,** e1000032 (2009)

When neuroscientists measure parts of nervous systems, they do so statistically, pooling data about the activity of many neurons at once. But researchers based in Cambridge, Massachusetts, have painstakingly mapped every neuron involved in innervating six mouse interscutularis muscles — muscles that allow mammals to wiggle their ears.

Jeff Lichtman of Harvard University and his colleagues used these six 'connectomes' to compare the innervation of paired tissues on the left and right sides of the same creatures. The wiring was strikingly different, underscoring the flexible nature of mammalian neural development. Many of the neurons were also 25% longer than required to form the connections that they did. That is odd because nerve cells are metabolically expensive.

An optical physicist sees beyond fluorescent labels.

Many a molecular biologist likes to watch molecules move around inside living cells, particularly in real time. The job is usually done by tethering a fluorescent tag to interesting biological molecules and following their movements by means of the tag's glow. But fluorescent tags are often bigger than the molecules they label, so frequently perturb their movements. Better to watch intracellular dramas without millstones around the actors' necks. But how?

A twist on 'Raman scattering' may hold the answer. Normally, when a laser is shone at a molecule, the molecule scatters most of the light at the same frequency at which it was emitted by the laser. A tiny amount — Raman scattered light — is scattered at different frequencies. These frequencies indicate the chemical bonds in the molecule, and can thus identify it as a fingerprint identifies a person. If only Raman signals were stronger, they would be suitable for real-time microscopy on a molecular scale.

A second laser provides the twist — and the necessary amplification. Sunney Xie of Harvard University and his colleagues have found that another laser can enhance the contrast of an image, improving the sensitivity over previous studies by four orders of magnitude (C. W. Freudiger *et al. Science* **322,** 1857–1861; 2008). For this to work, the two lasers must coincide on the sample, and the difference in their frequencies must exactly match that of a specific molecular vibration of a certain chemical bond in the sample. The background noise is eliminated and the signal is amplified.

This method is both versatile and powerful; the authors used it to observe the uptake of omega-3 fatty acids by human lung-cancer cells and the changing distribution of two drugs as they were absorbed by mouse skin. I think this could spur the development of tag-free molecular movie machines for all.

Discuss this paper at **http://blogs. nature.com/nature/journalclub**

NEWS

Antarctica's impossible peaks come into view

Major polar project yields panorama of hidden mountains.

A mysteriously rugged mountain range hidden beneath east Antarctica's massive ice sheet has been revealed in all its topographic glory by an international research team. The findings may cause geologists to rethink their ideas about the continent's history.

The mountains, named the Gamburtsev range after Russian seismologist Grigoriy Gamburtsev, were discovered during the International Geophysical Year in 1957–58, but the peaks were little more than points on a map to geoscientists. In the International Polar Year (2007–09), which concludes at the beginning of March (see News Features, starting on page 1072), an expedition surveyed the Gamburtsevs in December 2008 and January 2009 to produce the first detailed portrait of the buried range. Radar, gravity and other data from roughly 130 airplane flights reveal a jagged landscape dotted with rivers and lakes, hundreds of metres beneath the icy surface.

"They are incredibly rough mountains — they look like alligators' teeth," says US team co-leader Robin Bell of the Lamont-Doherty Earth Observatory in Palisades, New York. "It's a breathtaking landscape."

The rugged topography only

deepens the puzzle of how the Gamburtsevs formed, says Fausto Ferraccioli, UK team leader from the British Antarctic Survey in Cambridge, UK. The range sits in the middle of an ancient chunk of crust, where geologists had thought there had been no major tectonic activity for at least 540 million years. How long it's been there is a really good question," says John Goodge, a specialist in Antarctic geology at the University of Minnesota in Duluth, who was not involved in the project. If the range is indeed at least 540 million years old, then it is remarkably well-preserved for its age.

Another possibility is that the Gamburtsevs formed much more recently, as a result of volcanic activity. But preliminary results from the expedition don't show the large magnetic anomalies that would typically be expected when flying over volcanic terrain, says Ferraccioli. And recent work looking at sediments that spilled off the Gamburtsevs, before the ice sheet formed on top, also suggests that the range is not volcanic in origin (T. van de Flierdt *et al. Geophys. Res. Lett.* **35**, L21303; 2008).

Ian Dalziel, a geologist at the University of Texas in Austin, says there's no way to know exactly how old the Gamburtsevs are without drilling into them. To him, the ruggedness of the range is "a surprising result", with no convincing explanation so far. East Antarctica's ice sheet is thought to have formed roughly 35 million years ago, possibly nucleating atop the peaks of the Gamburtsevs.

The Antarctica's Gamburtsev Province Project (AGAP) was one of the major polar initiatives in the most recent Antarctic field season, involving researchers and support staff not only from the United States and Britain but also from Australia, Germany, China and Japan. The teams split into two camps; the US group worked from the south camp closer to the South Pole, whereas the UK team worked from the north camp on the other side of the Gamburtsevs (see map). Temperatures at both sites averaged –30 °C; and bad weather at the south camp, at 3,500 metres
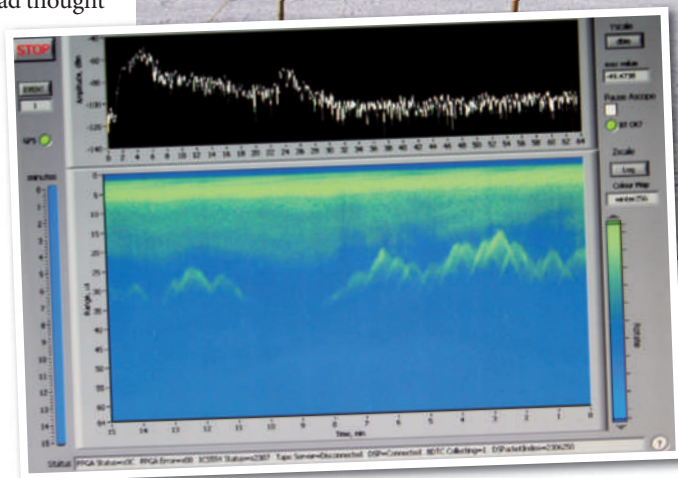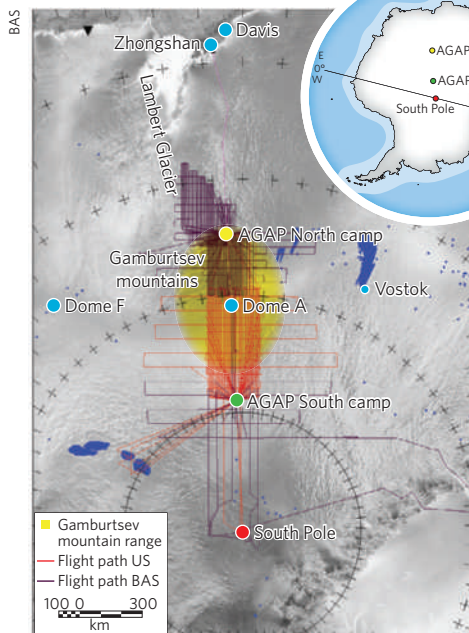

altitude, kept its two Twin Otter survey planes grounded for days at a time.

Eventually, the teams did manage to gather data while flying a total of 120,000 kilometres back and forth across the ice sheet in a dense grid; the US portion alone covered a region the size of California. The coverage is by far the most detailed look yet at the Gamburtsevs, although a Chinese expedition in the 2007–08 season gathered a line of radar data as it made its way from the coastal Zhongshan station to a new base China is constructing at Dome Argus (Dome A).

At their highest, the Gamburtsevs come within about 500 metres of the ice sheet's surface, says Bell. At the base of the mountains, the tremendous pressure of the overlying ice sheet and relatively warm temperatures has caused the ice to melt, creating lakes and rivers that can be seen in the radar data. As rivers cut ever deeper into the ground, the peaks could accordingly have been uplifted even higher, creating the dramatic topography seen today, says Ferraccioli. The Gamburtsev valleys are up to 700 metres deep from valley to peak, says Bell.

The data may also help ice-core specialists who are looking for a site to drill the next deep Antarctic core. Around Dome A — one primary target for drilling — the ice layers at

Map showing: Zhongshan, Davis, Lambert Glacier, AGAP North camp, Gamburtsev mountains, Dome F, Dome A, Vostok, AGAP South camp, South Pole. Inset: AGAP North, AGAP South, South Pole.

Legend:
- Gamburtsev mountain range
- Flight path US
- Flight path BAS
- 100 0 300 km

An international expedition has revealed jagged mountain ranges (left) under the Antarctic ice.

depth are "quite contorted", says Ferraccioli. That's not good news for those looking to drill unaltered ice layers to search for gas bubbles and other climate records from the past.

Meanwhile, AGAP's seismology team has deployed 24 seismological stations on the ice sheet above the Gamburtsevs. "This is the first time that there's been an array of stations in the middle of Antarctica that have run successfully all year long," says Douglas Wiens of Washington University in St Louis, Missouri, the team's co-leader. The stations have been gathering data on the subsurface structure by measuring seismic waves from large earthquakes, such as the Szechuan quake of May 2008, which pass through the Gamburtsevs on their way to the top of the ice sheet.

The seismic work was one of the few projects not affected by the funding shortages that have plagued polar research projects in recent years. The unrelated POLENET project, for instance, had to dramatically scale back the number of seismic and global positioning system stations it planned to install around Antarctica in the most recent field season. And in AGAP, the airborne survey had to cut back on its planned number of flights by about 20%, says Bell.

Still, she says, the Gamburtsev project wouldn't have happened any other way. "The framework of the IPY," she says, "really let us launch a major expedition in the most difficult place to work." ■

Alexandra Witze
See Editorial, page 1057.

Harbour threat for coelacanths

A population of coelacanths — the rare deepwater fish once thought to have been extinct for millions of years — could be at risk if a proposed harbour project in Tanzania goes ahead, conservation experts warn.

A port development plan, commissioned by the Tanzania Ports Authority, recommends building a deep-water harbour in the shallow Mwambani Bay south of Tanga, a coastal town close to the Kenyan border. The Tanzanian government hopes that the new port, designed to accommodate large container ships, will stimulate economic development in the region.

Plans for such a harbour date back to 1977, but authorities are now moving forwards fast, having designated around 80 hectares of land from which they will have to evict local residents. A Chinese–Tanzanian agreement on economic and technological cooperation, signed last week by the Chinese president Hu Jintao during a state visit to Dar es Salaam, could lend support to the US$164-million project.

Opponents complain that the local business community, conservation organizations and residents of Mwambani Bay have not been involved in the planning. Local seafood exporter Eric Allard, who trained in Florida as an oceanographer, says that the existing port in Tanga operates at less than 50% of capacity. "There is no need whatsoever for a new harbour," he says.

However, environmental activists say that it might be related to plans to build a soda-ash (sodium carbonate) extraction plant 500 kilometres away on the shores of Lake Natron. The plant would be able to produce up to 1 million tonnes of soda ash annually, and the planned new harbour could provide the capability to ship it.

Environmentalists are concerned about potential damage to the local ecosystems. A 2008 technical assessment by the Tanzania Natural Resource Forum (TNRF), a non-governmental environmental organization based in Arusha, concluded that building a harbour in Mwambani Bay "will probably wipe out the local coelacanth population".

Coelacanths (*Latimeria chalumnae*) were 'rediscovered' in 1938 when a specimen was caught off South Africa. Off Tanzania, the first coelacanth was caught in 2003. Since then, 50 or so catches of coelacanths have been reported by Tanzanian fishermen. Scientists think that the fish are under pressure from fishing with dynamite and the use of deep-set shark nets. A population off the Comoros Islands in the Indian Ocean is now stable thanks to conservation measures initiated almost 20 years ago, says Hans Fricke, a marine biologist with the Leibniz Institute of Marine Sciences in Kiel, Germany. The Tanga population probably consists of strays from the Comoros population, he says.

The species' small gene pool offers a unique window into the history of life on Earth, says Mike Bruton, a consultant with architecture and design company MTE Studios in Cape Town and former director of the J. L. B. Smith Institute of Ichthyology in Grahamstown, South Africa, named after the man who identified the coelacanth in 1938.

Coelacanths are found in Tanzanian waters.

"The international community should take very seriously anything that threatens such an important species," says Bruton. "Building a harbour right in the middle of a proclaimed preserve will substantially harm underwater habitats."

Conservation experts are meeting this week in Tanga to discuss options for a Marine Protected Area for coelacanths, which the Tanzanian government promised to create in 2006. Tanzanian law requires that an environmental and social impact assessment be carried out before construction starts. But no such exercise is under way. Tanzania's National Environmental Management Council recently told the TNRF that it was not involved in the planning, nor was it aware of any impact assessment. ■

Quirin Schiermeier

Drug patent plan gets mixed reviews

GlaxoSmithKline's bid to tackle neglected diseases receives a muted response from the rest of the industry.

Proposals to radically change the way the drug industry approaches neglected tropical diseases have prompted intense debate within the sector.

Andrew Witty, chief executive of GlaxoSmithKline (GSK), outlined the suggestions in a speech on 13 February at Harvard Medical School in Boston, Massachusetts. He committed GSK — the world's second-largest pharmaceutical company by sales — to sharing some of its patents to boost research into neglected diseases, and to making its drugs available more cheaply in the very poorest countries.

He also offered to open up GSK's research centre for neglected diseases at Tres Cantos in Spain to other researchers, companies and governments. The aim would be to foster a global public–private network to supplant the present fragmented research efforts on the most neglected diseases, such as sleeping sickness, visceral leishmaniasis and dengue fever.

Reactions from scientists and public-health experts have ranged from wildly enthusiastic to deeply sceptical. But it is undoubtedly a pioneering move, particularly for a firm that just a decade ago joined 38 other drug companies in suing the South African government to try to stop it making cheap anti-HIV drugs.

"That was the low-water mark for multinationals and global health," says Peter Singer, an expert in public health at the University of Toronto in Canada. "This announcement may not yet be the high-water mark, but it is incontrovertible evidence that the tide has turned. It sends a clear signal that GSK wants to be part of the solution and not part of the problem."

The most innovative aspect of GSK's proposal is the creation of a 'patent pool' for drugs and manufacturing processes related to neglected tropical diseases. Researchers and companies, including manufacturers of generic drugs, would be able to license participants' patents from the pool for free to develop new treatments for neglected diseases in the world's 50 least-developed countries (LDCs).

"A patent pool giving access to molecule libraries and information from different groups

Making plans: Andrew Witty.

is something we have been asking for for years," says Bernard Pécoul, head of the Drugs for Neglected Diseases Initiative based in Geneva, Switzerland. But its success depends on what companies are willing to put in, he adds. "If they put garbage in, we will get garbage out."

Although the pool would be oriented towards drug discovery, it could also provide access to promising drug candidates, existing drugs and formulation technologies for commercialization.

"It's a fantastic step forward," says Mary Moran, director of health policy at the George Institute for International Health in Sydney, Australia. Drug companies usually fight to defend their patent portfolios, she says, refusing to allow their free use for good causes such as tackling neglected diseases. As the patents in this sector have little monetary value, this amounts to "two bald men fighting over a comb", she observes wryly.

Industrial unease

But several other drug companies contacted by *Nature* were lukewarm about the idea. French firm Sanofi-aventis says that the proposals are "too vague" to comment on, and Swiss-based Novartis "does not consider intellectual property as an obstacle to access to medicines", according to company representatives.

In principle, Bayer in Germany is not against the idea of pooling intellectual property, says the company's spokeswoman, Denise Rennmann. But it would prefer an industry-wide approach to be agreed multilaterally, she adds — something that the World Health Organization and the International Federation of Pharmaceutical Manufacturers are already discussing. Many companies highlighted their own efforts to improve access to drugs for neglected diseases (see 'Opening access').

The patent-pool proposal has also sparked controversy because it excludes GSK's HIV patents, as the company feels that there is already enough research in this area. That is hotly contested by Michelle Childs, director of policy and advocacy at Médecins Sans Frontières' Campaign for Access to Essential

---

OPENING ACCESS

A selection of other drug firms' efforts on neglected tropical diseases.

Pfizer
● Donates medicines through individual deals with governments and non-governmental organizations.
● Mobilize Against Malaria initiative in Ghana, Kenya and Senegal to improve access to malaria treatment.
● International Trachoma Initiative has donated 54 million antibiotic treatments to patients in 13 countries (see *Nature* **457**, 772–773; 2009).

Hoffmann-La Roche
● Expert assistance given to the Medicines for Malaria Venture, a non-profit organization aimed at discovering malaria drugs.
● Donated to the Brazilian government all rights and technology to manufacture benznidazole for the treatment of Chagas' disease.
● Technology-transfer initiative helps developing countries produce their own generic HIV medicines.

Novartis
● Provides not-for-profit treatments for dengue fever, tuberculosis and malaria in countries where endemic.
● Novartis Institute for Tropical Diseases in Singapore established in 2002 as a public–private partnership with the Singapore Economic Development Board.
● Novartis Vaccines Institute for Global Health in Siena, Italy, established in 2007 to develop vaccines for neglected diseases.

Sanofi-aventis
● Sells key medicines at cost to governments and aid agencies.
● In 2007 spent €14 million (US$18 million) on improving access to medicines in developing countries; €4 million on neglected tropical diseases; and more than €20 million on R&D for malaria, leishmaniasis and tuberculosis.

Bayer
● Donates drugs for sleeping sickness and Chagas' disease to the World Health Organization.
● Donates antibiotics for tuberculosis to the Global Alliance for TB Drug Development.
● Bayer CropScience donates mosquito nets and insecticides in key countries. **N.G.**

T. MELVILLE/REUTERS

Medicines, who says that there is still a great need for new antiretroviral combinations and formulations for children.

However, GSK's HIV patents could yet dive into a different pool. UNITAID, an international organization that negotiates lower drug prices, hopes to launch a patent pool by the end of this year that would allow companies to license their AIDS drugs in return for royalties. Ellen 't Hoen, UNITAID's senior adviser on intellectual property, says that although no companies have yet signed up, several have expressed interest, including GSK.

There is a sound incentive for companies to join such patent pools. Countries with large populations of sick patients are increasingly likely to use their rights under TRIPS, the World Trade Organization's intellectual-property agreement, to issue compulsory licences, allowing them to produce generic copies of patented drugs at low royalties.

Western drug companies fret that such licences could limit their opportunities in what they see as their biggest future growth markets: the well-off elites in emerging economies such as India, China or Brazil. Preserving these markets may explain why GSK has limited its proposals to the LDCs, says Brook Baker, an expert on health and human rights at Northeastern University in Boston.

Witty also announced that GSK would cut the prices of all its medicines in the LDCs to no more than 25% of their prices in the richest countries. But the focus on the LDCs means that these cuts, and the patent pool, will not benefit the vast number of poverty-stricken people living in middle-income countries, says Baker. According to the World Bank, more people live on less than $2 per day in India than in all of sub-Saharan Africa.

Despite the cuts, GSK's drugs will still be unaffordable to most poor people, he adds. And if generic producers can only use the patent pool to produce drugs for the LDCs, they will not be able to get the economies of scale they need to sell drugs cheaply enough.

"When Glaxo and others announced at the start of the decade that they would cut the cost of HIV drugs by 70%, initially we thought 'this is a big deal'," says Childs. "But they were still unaffordable. It was only when generic manufactures brought prices down 99% that we saw wide access to drugs." ■

Declan Butler
Additional reporting by Natasha Gilbert
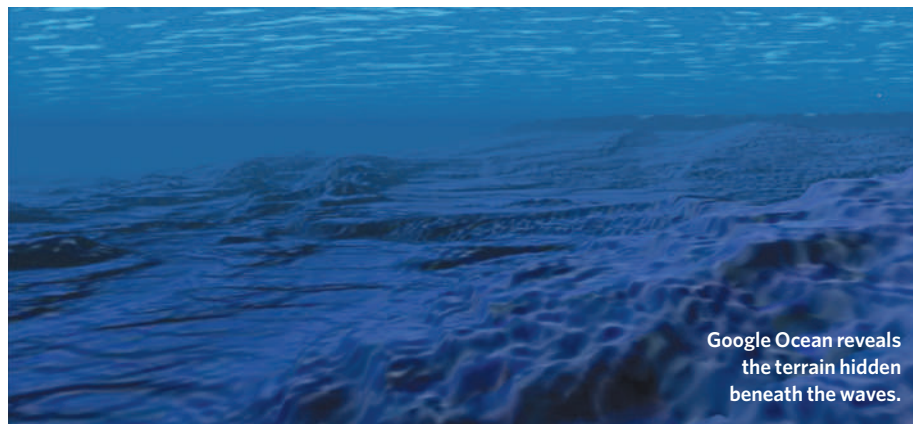
Google Ocean reveals the terrain hidden beneath the waves.

Seabed images create waves

Fly to coordinates N0 E94.75 in Google Ocean, the visualization program launched this month that renders marine environments in unprecedented detail, and you'll see something that's definitely not in the real ocean: the initials 'DTS/SIO'. This is the mark of geophysicist David T. Sandwell of the Scripps Institution of Oceanography in La Jolla, part of the University of California, San Diego (UCSD), who uses the initials here and elsewhere to mark vast areas of the ocean that remain uncharted.

The initials are also a reminder of the people and institutions behind the data — and how they can lose out on financial gain as public assets enter the corporate world. More than a decade ago, Sandwell and Walter Smith of the National Oceanic and Atmospheric Administration in Silver Spring, Maryland, developed a way to convert gravity and ocean-depth data from satellite missions and research cruises into a bathymetric map of the sea floor. Google Ocean incorporates that method — which Sandwell copyrighted for UCSD — without compensation.

Many researchers have praised Google's visualizations such as Google Earth and Google Mars. "Scientists want to see their data published and used by the world," says Steve Miller, project manager for Google Ocean at the California-based company headquartered in Mountain View. The programs incorporate public-domain data from non-profit organizations and agencies such as NASA.

But some question whether institutions — whose researchers often develop sophisticated ways of using such data — are losing out. "The marketplace is taking the technology and the university isn't doing squat," says Gideon Markman, a management

"The university isn't doing squat."

scholar at Colorado State University in Fort Collins. Jeff Chester, a social scientist who directs the Center for Digital Democracy in Washington DC, goes further: "In this case, it looks like the university was digitally fleeced."

Miller says that Google talked last summer with UCSD and Sandwell about incorporating the technology. "There were some legal questions," he says. "But we are comfortable." And William Decker, UCSD's technology-transfer official on the case, says, "we followed the process we always do to get appropriate fair value".

Sandwell wants other researchers to use his bathymetric map, and asks commercial providers to contact him before use. Oil and gas companies such as ExxonMobil have done so and funded his research. But Google didn't contact Sandwell when it first started using an earlier version of the bathymetry in 2005, and he raised the issue with the company last year as it was working on its oceans system.

Tony Haymet, the Scripps's director who led discussions with Google, acknowledges that there were no negotiations for compensation. Asked why the university didn't push to receive anything when state universities are facing major financial difficulties, Haymet responded: "That's a good question."

Last August, the US National Geophysical Data Center in Boulder, Colorado, pulled a version of Sandwell's technology from its website in part because of the copyright issue.

So far, all that Sandwell has received in return for his work is a year's worth of Google Earth Pro software — normal cost US$400 — for his classes. He is now applying to a Google foundation for funds to support a postdoctoral research position in oceanography. ■

Rex Dalton

Hong Kong evolution curriculum row

A debate over the teaching of intelligent design in schools is raging in the unlikely battleground of Hong Kong. A fiery topic in the United States, intelligent design — which holds that a designer has shaped life on Earth — has gone almost unnoticed in Asia.

The controversy is affecting all levels of education. In the latest episode, a dean of science at one of Hong Kong's premier universities backed out of a radio show earlier this month after finding he was being set up to debate with an advocate of intelligent design within his faculty.

Hong Kong is in the middle of reforming its school system, and the Education Bureau has issued a series of guidelines for all levels of education to go into effect in September 2009. Within the 126-page guidelines for teaching biology at the secondary level, the section on 'Genetics and Evolution' includes the sentence: "In addition to Darwin's theory, students are encouraged to explore other explanations for evolution and the origins of life, to help illustrate the dynamic nature of scientific knowledge." The guidelines were prepared by the Curriculum Development Council and the Hong Kong Examinations and Assessment Authority.

Early this month, in the *South China Morning Post*, several senior scientists including astronomer Sun Kwok, freshwater ecologist David Dudgeon and geologist Jason Ali, all of the University of Hong Kong, called for the

Science faculty will not teach intelligent design.

sentence to be omitted. In the same article, a spokeswoman for the Education Bureau defended the teaching of alternative explanations, because science cannot "provide a complete answer" about the emergence of life.

There has been heated debate in the Hong Kong media since. On 13 February, the radio series *Backchat* broadcast a programme on 'Creationism versus Evolution'. Kwok, dean of the science faculty at the university, backed out of the programme when he found it had been

changed to a debate format at the last minute, noting that he is not a specialist in evolution and thus "cannot counter detailed arguments regarding evolution versus creation".

Chris Beling, a solid-state physicist and associate professor in the University of Hong Kong department of physics, debated with two other panellists and took the opportunity to "mildly criticize" the science faculty for refusing his request to continue a course on the origin of the Universe that included a section on intelligent-design thinking. Instead, he has met in private with students to discuss intelligent design.

The depth of interest in intelligent design in Hong Kong is rare for Asia, says Chung-I Wu, an evolutionary biologist at the University of Chicago and a visiting director at the Institute of Genomics in Beijing. "Each of us card-carrying evolutionists has encountered issues of intelligent design in Taiwan or China," he says, "but the challenge is never serious enough to leave even a name or organization that we can recall." Beling says that the presence of many missionary schools in Hong Kong may explain why intelligent design has taken root there.

For now, Kwok and his university are unlikely to back down. "The position of the faculty of science of the University of Hong Kong is that pseudoscience subjects such as intelligent design, astrology and UFO studies have no place in our science curriculum," says Kwok. ■
David Cyranoski

M. GRANT/ALAMY

Satellite to monitor carbon sinks sinks

A mishap shortly after launch led to the loss of the Orbiting Carbon Observatory (OCO), a NASA environmental monitoring satellite, on 24 February. The satellite crashed into the ocean off Antarctica.

"It's a major setback," says Paul Palmer, a climate scientist at the University of Edinburgh, UK, who is part of the OCO science team. It will be particularly devastating for the tight-knit group of scientists and engineers who have devoted much of the past decade to the project. "These guys have sweated OCO for seven or eight years," he says.

OCO, which cost US$273 million, was designed to measure carbon dioxide levels at various depths in the atmosphere with high enough precision to allow the sources and sinks of the gas to be assessed. It would have provided researchers with a comparatively

high-resolution global picture of how carbon dioxide is absorbed by oceans, forests and other components of the Earth system.

The satellite was launched aboard a Taurus XL rocket at 4:55 a.m. local time from Vandenberg Air Force Base in California. Shortly after launch, controllers announced that the protective clamshell 'fairing' that surrounded the satellite had failed to separate properly.

The Taurus was built by Orbital Sciences Corporation of Dulles, Virginia; the company was also the main contractor on the spacecraft itself. This was the eighth launch of a Taurus rocket and the second failure. An investigating board was being set up as *Nature* went to press. The next payload booked for launch on a Taurus is another NASA satellite, Glory, which will study clouds and aerosols. It will not now be

launched until OCO's loss is understood.

OCO was one of two carbon dioxide-monitoring satellites to launch this year. In late January, the Japanese Aerospace Exploration Agency launched the Greenhouse Gases Observing Satellite (GOSAT), which will monitor levels of methane, water vapour and carbon dioxide. That satellite seems to be functioning normally.

The two satellites would have been complementary to each other. Whereas GOSAT provides high spectral resolution of gases in the atmosphere, OCO would have provided detailed spatial resolution of carbon dioxide. In addition, the two satellites would have allowed scientists to check their data against one another. "Having two satellites was going to make life a lot easier," says Palmer. ■
Geoff Brumfiel

Four arrests in Californian animal-rights activism

An international investigation has led to arrests of four animal-rights activists suspected of harassing researchers working at University of California campuses.

Agents of the Federal Bureau of Investigation seized two of the suspects on 19 February in Charlotte, North Carolina, and the other two in Oakland, California, the following day.

The four are suspects in two incidents: an alleged trespassing involving a researcher from the Berkeley campus in 2007 and a home invasion at the residence of a professor at the Santa Cruz campus in 2008. Federal authorities are continuing an investigation of firebomb attacks on homes of researchers from the Los Angeles and Santa Cruz campuses.

Commercial biofuels plant planned for Florida

London-based oil giant BP is teaming up with the biotech firm Verenium in Cambridge, Massachusetts, to build a biofuels plant in Highlands County, Florida, that will produce 136 million litres of ethanol a year from non-food crops such as grasses, the companies announced on 18 February.

The commercial-scale plant, which will begin production in 2012, is one of the largest cellulosic-biofuel initiatives to be announced, and will cost up to US$300 million to build. Around 8,000 hectares will be planted as feedstock.

The partners — who have each put $22.5 million into the joint venture to date — say that they plan to build other full-scale cellulosic-biofuel facilities in the Gulf Coast region of the United States.

Iran's uranium stockpile reaches weapons potential

Iran's stock of low-enriched uranium hexafluoride would, if enriched to weapons-grade purity, be enough for a nuclear bomb,

Europa next outer-planet mission

NASA and the European Space Agency (ESA) have chosen Jupiter's moons as the destination for their upcoming joint outer-planet flagship mission. A trip to Saturn's moon Titan may yet go ahead but needs further investigation, the agencies announced on 18 February.

The Europa Jupiter System Mission will launch two orbiters, one built by NASA and the other by ESA, in 2020, with a scheduled arrival time in the Jupiter system of 2026. The NASA orbiter will study the icy shell of Jupiter's moon Europa — which may hold an ocean capable of supporting life — and will cost US$2.5 billion to $3 billion, says Jim Green, director of NASA's planetary-science division.

The European part of the mission, called Laplace, will investigate Ganymede, the largest moon in the Solar System. It must still compete with two other missions to secure funding.

A longer version of this story can be found at http://tinyurl.com/bovl2j.
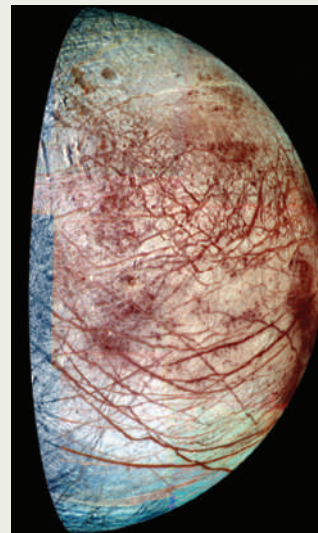
NASA/JPL/UNIV. ARIZONA

according to the International Atomic Energy Agency (IAEA).

The United Nations' nuclear watchdog reported on 19 February that the country now has more than a tonne of low-enriched uranium. That is much more than the agency's November 2008 estimate of 660 kilograms, suggesting that earlier stocks were misreported. Much of the excess was discovered during the agency's annual inventory of materials at Iran's uranium-enrichment facility at Natanz.

To be used in weapons, uranium must be enriched to around 90% pure uranium-235. Iran has so far enriched its uranium to just 3.49%, or reactor-grade, the IAEA says.
A longer version of this story can be found at http://tinyurl.com/cn8snu.

deCode gets marketing go-ahead in California

Eight months after being targeted by California regulators with a 'cease and desist' letter, deCode Genetics, a biotech company based in Reykjavik, last week received the equivalent of the state's blessing: a clinical laboratory licence.

deCode's in-house laboratory has passed muster with state regulators, and the company can begin marketing genetic tests in California. Last June's letter threatened deCode with penalties of up to US$10,000 a day for marketing to California residents (see *Nature* 453, 1148–1149; 2008).

Two other high-profile firms — 23andMe and Navigenics — that received the letter were granted licences in August.

Kári Stefánsson, deCode's chief executive, says that the firm is awaiting a similar licence from the state of New York.

Forensic science badly lacking in the United States

The US forensic-science system is fragmented, underfunded and poorly supported by research, according to a report issued by the US National Research Council on 18 February.

The judiciary system relies on forensic science to link evidence to a suspect, but only nuclear-DNA analysis has been shown to make such connections reliably, the report says. The government should create a federal agency to support research on forensic methods and establish ways to quantify uncertainty, says the council. Forensic-science practitioners should undergo mandatory certification, and laboratories should be separated from law-enforcement and prosecutors' offices.

Television shows such as CSI give juries unrealistic expectations of forensic evidence.

The US Senate judiciary committee will review the report and decide on the necessary hearings and legislation, a statement from chairman Patrick Leahy (Democrat, Vermont) says. Senator and ranking Republican of the committee Arlen Specter (Pennsylvania) says it is "incumbent on Congress to pursue the report's findings".

CBS/J. BRUCKHEIMER TELEVISION/TOUCHSTONE TELEVISION/RONALD GRANT ARCHIVE

IN FROM THE COLD

As scientists celebrate the end of the International Polar Year, they see causes for concern on the frozen horizon, reports **Quirin Schiermeier**.

It is not often that inaction earns a spot in history, but a French research team has managed to write itself into the annals of polar exploration by simply waiting — albeit in a perilous spot. The record-breaking feat happened on board a privately owned French schooner, *Tara*, which was frozen solid in thick Arctic pack ice. For 17 months, the ship and its 10-member scientific crew passively rode with the ice, and on 28 May 2007, the ship came closer to the geographic North Pole than any ship ever before. At a latitude of 88 degrees and 32 minutes north, the *Tara* was just 160 kilometres away from the pole.

That was not the only remarkable aspect about this expedition, one of many newsworthy cruises under the flag of the fourth International Polar Year (IPY), which runs from March 2007 to March 2009. Scientists on board the *Tara* had planned to float with the pack ice for two full Arctic winters, but their trip ended almost 9 months early because of extremely rapid ice movement — an indicator of just how fast the Arctic is changing.

"We realized after the first winter that we had drifted much faster than predicted," says Jean-Claude Gascard, an oceanographer at the Pierre and Marie Curie University in Paris and chief scientist of the expedition. The early exit from the ice came at an inopportune time — it halted data collection much earlier than planned and forced the crew to make a dangerous run towards the island of Svalbard. "Sailing back to Svalbard in the middle of the Arctic winter was a pretty precarious thing to do," says Gascard.

Its success in the face of uncertainty made the *Tara* expedition "symbolic for the spirit of the IPY", says David Carlson, an oceanographer who heads the IPY programme office in Cambridge, UK. "To me it was its most iconic moment."

Global challenge

This IPY has seen some 50,000 scientists, students, technicians and crew participate in expeditions and scientific programmes focusing on the Arctic and the Antarctic. Its 170 collaborative international projects have covered disciplines from ecology to astronomy, from the social sciences to human and animal health. In the Arctic, indigenous people and their traditional knowledge became part of the science wherever possible, sometimes inspiring it in unanticipated ways (see page 1077).

It might seem that, as so often in the past, science reigns supreme at the planet's poles. But as climate change opens up vast parts of the Arctic to commerce, nations are starting to exert their influence in the region more purposefully, and long-simmering political tensions might soon boil over.

Scientists on the French schooner *Tara* spent 17 months locked in the Arctic pack ice.

F. LATREILLE

The current IPY extends a tradition that reaches back almost 130 years. Each of the three previous polar years — in 1882–83, 1932–33 and 1957–58 — kicked off a new era in polar science. In the most recent IPY, the main scientific challenge was to explore the most remote and hostile spots on the globe. It was an era when daring graduate students could still get their names on peaks never before seen by human eyes. The focus over the past two years, by way of contrast, has been on acronym-bearing studies that encompass the physical, biological and social components of the polar environments.

As in other areas of science, many polar researchers now take a systems approach, examining, for example, interactions between glaciers, ocean circulation, sea ice, algae and seal health. These and many other parts of the polar system have been studied before. But the scope of activities carried out during this IPY would have been unthinkable without the technological advances of recent decades, including such tools as satellites, laptop computers and high-tech clothing.

Most importantly, this IPY has proceeded amid galloping climate change, which has already altered the Arctic in dramatic ways and is increasingly making itself apparent around Antarctica. As the *Tara*'s speedy voyage illustrates, the scale of the polar changes dwarfs what many researchers would have thought possible even just a few years ago. "There is no final word yet as to what has made the ice drift so fast," say Gascard. "But it is pretty clear that changing wind patterns and changes in the thickness, structure and concentration of ice have all played a part."

The speed of the expedition was hardly a coincidence, he adds. On an ice floe somewhat farther south, a Russian team camped out over winter and moved even faster than the *Tara*. In fact, the Russians drifted faster than any other Arctic ice station established by that country over the past 34 years. Compared with the era of the Norwegian explorer Fridtjof Nansen, who took a roughly similar cross-Arctic trip on board the *Fram* 115 years ago, sea ice seems to be drifting almost twice as quickly.

The fast movement means that sea ice is more likely to make it to the Atlantic, and that is one of the reasons for its dwindling extent: the ice cover reached a record low in September 2007, followed by only minimal recovery last summer. The Canadian icebreaker *Amundsen* sailed for a full year along Canada's entire Arctic coastline in flaw leads, semi-permanent patches and channels of open water between land-fast ice and the

J.-G. WINTHER

A Norwegian–US team takes a long drive across Antarctica to get climate and geological data.

drifting pack. That expedition conducted the most extensive survey ever of how retreating sea ice affects marine ecosystems (see *Nature* **454**, 266–269; 2008). The 400-person, US$40-million expedition, one of the largest IPY projects, yielded a plethora of data, samples and *in situ* observations. Analysing the results will keep scientists busy for years.

Lone wolves

The scientists involved in the *Amundsen* project and almost all those working in other IPY ventures made it through their expeditions safely, although a helicopter crash in March last year claimed the lives of a pilot and a technician near Germany's Antarctic station Neumayer II. And that was not the only setback for the polar-year programme, which also ran into rough political waters involving tensions between the East and West. The *Tara* expedition, for example, almost had to be cancelled when Russian authorities denied necessary logistical support, preventing the crew from leaving the Siberian port of Tiksi for two weeks. At almost the last possible minute, the authorities finally gave the green light for a Russian icebreaker to accompany the *Tara*, allowing the schooner to set sail.

Many scientists had hoped that this IPY would open doors on a new era of Arctic and Antarctic collaboration with Russia, which has in the past conducted most of its polar science activities in isolation. These hopes, however, were not fully realized. The bulk of Russia's polar-science programme during this IPY

> "Scientists can, and I believe should, intervene to influence the path taken in the years to come."
> — Oran Young

consisted of activities conceived with a national focus. Russia's planting of a flag on the Arctic seabed in 2007 symbolized its claims to a large chunk of the Arctic Ocean and reflected its national interests in the region (see *Nature* **448**, 520–521; 2007).

"We are proud of our long tradition in polar science and exploration," says Denis Moiseev, deputy director of the Murmansk Marine Biological Institute of the Russian Academy of Sciences' Kola Science Centre. "We are going on expeditions every year to do ecosystems research in polar regions, almost independent of whether or not there happens to be an International Polar Year."

But the IPY, to which the Russian government contributed some US$10 million, has at least facilitated some joint undertakings with Western countries. The best example is the International Siberian Shelf Study, which looked at climate change, permafrost and ocean chemistry along Russia's vast northern coastlines and involved some 30 scientists from four other nations. The expedition, which travelled aboard the Russian research ship *Jacob Smirnitskyi*, came about in part because of personal contacts. Its chief scientist, Igor Semiletov, bridges two countries, as a senior researcher at the Pacific Oceanological Institute in Vladivostok and a visiting scientist at the University of Alaska in Fairbanks.

Russia was one of four nations — alongside Canada, Norway and the United States — to be involved in more than half of the IPY projects. For the first time ever a Westerner was invited to Russia's drifting ice station (see *Nature* doi:10.1038/news.2008.956; 2008). Furthermore a Russian–French–Italian team undertook one of the

NOTES FROM THE POLES

The International Polar Year (IPY) has covered two full annual cycles at both poles. Consisting of 170 projects, it has involved more than 60 countries and cost about US$1.2 billion. Here are some snapshots.

ARCTIC

1.3.07–1.3.09

When it comes to climate change, the Arctic is regarded as the planet's early warning system: the region is expected to respond faster and more severely than the rest of Earth. The IPY effort took the pulse of the Arctic.

CONTINUOUS PERMAFROST

North Pole

Arctic Circle

Inuit research
work during the polar year.
The icy cover on the Arctic Ocean is fast disappearing during summer. One programme sought to tap the knowledge of Inuit and other residents about it. Another project in Canada used an icebreaker to visit over 30 remote communities to collect health data.

Permafrost covers nearly one-fifth of the land area in the Northern Hemisphere, but global warming is thawing this frozen ground. At a site near Tiksi in northern Russia, a researcher measures the temperature 15 metres below the surface. Data from that site for the past 2 years show the annual cycle as well as an overall warming trend. Permafrost in northern Russia and Alaska has warmed 1–2° C since 1970.

-8.9 °C
-9.1 °C
-9.3 °C
-9.5 °C
2007   2008
Researcher: A. Kholodov
The rising temperature of the permafrost

Wish you were here?

Subject: Svalbard contamin
When measuring pollution, it pays to stay clean. French researchers wear special suits while sampling snow in Svalbard for mercury contamination.

FACSIMI
Just back from our collaboration with the International Heliophysical Year studying the auroras to investigate connections between the Sun and Earth's atmosphere.
Did you get my postcard?
See you soon.

ANTARCTIC

1.3.07–1.3.09

For much of the past century, Antarctica has been the playground of adventurers and scientists. This polar year was no exception. Dozens of projects focused on exploring and monitoring the southern continent, which is a wild card in the realm of climate change.

The team en route to Dome A, the site of the new Chinese Antarctic base.

Census areas marked in red

Troll station
Pole of inaccessibility
South Pole   *Norway–US traverse*

Megaleledone setebos

Census of Marine Life find
The Census of Antarctic Marine Life sent out an armada of ships, which discovered hundreds of new species. The survey showed that the initial cooling of Antarctica 35 million years ago sparked the evolution of new octopus species and their spread to other deep ocean basins.

Surveys also show that identical animals thrive at both poles.

Plane crashes in Antarctica
A plane carrying 6 researchers and 4 crew crashed during take-off on 20 December 2007. No one was seriously injured. The scientists on board had just installed GPS and seismic stations near the Antarctic coast, part of a project to track changes in the ice sheets and the bedrock below.

The Andrill site

ANDRILL PROJECT
- US$30 million
- More than 100 scientists
- 2 kilometres of rock recovered

SUMMARY:
Two boreholes into the sea floor near Ross Island, the deepest in Antarctica, recorded the history of that region over the past 20 million years.

BACKGROUND MAPS: H. AHLENIUS, UNEP/GRID-ARENDAL
ARCTIC MAP: Full data sources available at http://tinyurl.com/byss9f.
ANTARCTIC MAP: Full data sources available at http://tinyurl.com/aofzdc.

ARTWORK: N. SPENCER AND W. FERNANDES; TEXT: R. MONASTERSKY

G. WILTSIE/NAT. GEOGR./GETTY
V. MAMYKIN
ARCTIC-IMAGES/CORBIS
C. MOREL/OUR POLAR HERITAGE
J. DE POMEREU/IPF
M BARKLAGE
M. RAUSCHERT, UNIV. ALASKA FAIRBANKS, COML
R. HOPCROFT, UNIV. ALASKA FAIRBANKS, COML
ANDRILL SCIENCE MANAGEMENT OFFICE

Ice cores extracted from remote spots will reveal details about Antarctica's climate history.

<div style="text-align:right">S. TRONSTAD, NORWEGIAN POLAR INSTITUTE</div>

traverses of Antarctica during this IPY. And scientists from Norway were involved in smaller research projects with Russia on reindeer herding and human health in Siberia.

Warming in the Arctic, and the retreat of summertime sea ice, is opening up the region to interests such as mineral exploitation, shipping, fishing and tourism. Some researchers fear that the commercial potential could shift international interactions from mainly scientific collaboration to hard-nosed politics. Environmental groups such as Greenpeace have proposed a 50-year moratorium on all exploitation in the Arctic, but this is unlikely to gain much support.

The shift towards economic and geopolitical competition poses a new threat for vulnerable Arctic environments, which should prompt scientists to speak out, says Oran Young, an expert on international governance and environmental policy at the University of California, Santa Barbara. "Whither the Arctic? What's needed is science- and ecosystem-based management, rather than a race for Arctic resources with inevitable clashes of national interests," he says. "Scientists can, and I believe should, intervene to influence the path taken in the years to come."

Friction in the cold

It is too early to judge how IPY science can help improve the management of the poles, says Young. "It has certainly done no harm," he says. One way forward is for scientists to start applying the results of their research. "I would like to have seen more collaboration and more extensive dialogue between polar researchers and the wider environmental-change community. It is not enough to read each others' writings. We also need to bring together the scientific communities and policy-makers at venues and conferences," he says.

> "There's a lot we could celebrate. But what's on our radar, much more than celebration, is continuity."
> — Jan-Gunnar Winther

With climatic and economic stakes growing at the poles, political leaders around the world are devoting more attention than ever to the Arctic and Antarctic. A record number of 63 nations participated in this IPY. Countries such as China, Belgium, Poland and Spain each took part in more than 10% of IPY projects and have joined the previously small club of nations that drive polar science. A fair number of scientists from more unlikely nations, including Bermuda, Brazil, Kenya, Mongolia, Uzbekistan and Vietnam, have also been involved.

Despite that attention, funding was not spectacular. Overall, the participating nations spent some US$1.2 billion on IPY research, but only around one-third of that was money over and above typical annual investments in polar studies. Four nations — the United States, Canada, China and Norway — provided three-quarters of this additional money. But, says Carlson, even modest funding had a considerable effect in countries such as Brazil, Malaysia and Portugal, which had almost no money available for polar science before the IPY.

One-third of IPY projects took place in and around Antarctica. Two of the biggest projects were the largest-ever census of Antarctic marine life and a series of six coordinated inland traverses of the unpopulated continent. Researchers made the long-distance tractor trips, at high altitudes and extremely low temperatures, to measure the conditions and drill ice cores in areas from which no *in situ* data had ever been collected. Besides gathering basic information about the climate's history and the geology below the ice sheet, these expeditions also aimed to provide 'ground-truth' data to calibrate satellite recordings. By comparing data taken at the surface with satellite-based readings, researchers can more precisely track changes in the ice thickness and its movement, among other parameters.

The US–Norwegian traverse of east Antarctica took a round trip route from the Norwegian Troll station on the coast, to the South Pole and back. Along the way, it passed the notorious Pole of Inaccessibility, the spot in Antarctica farthest from the Southern Ocean, much more difficult to reach than the geographic South Pole. The team took the first-ever surface-based measurements of sub-glacial lakes in the area, which lie hidden beneath three to four kilometres of ice. "It had the air of true exploration, the way it was done in the early days," says Jan-Gunnar Winther, director of the Norwegian Polar Institute in Tromsø, who co-led the traverse with Mary Albert of the US Army's Cold Regions Research Engineering Lab in Hanover, New Hampshire.

"Such large Antarctic projects require an awful lot of international coordination in terms of logistics, science and funding," says Winther. "Without the IPY, many of the things we were able to do here just wouldn't have happened."

The global financial crisis came towards the end of the IPY and so did not cripple operations as much as it might have, had the timing been different. However, roaring fuel prices, which peaked last summer, did hit all polar operators. Some Antarctica programmes funded by the US National Science Foundation, such as a geophysical study of the Gamburtsev Mountains in the east of the continent, got less air support than planned, for example (see page 1062). It also meant delaying the deployment of seismic sensors for the Polar Earth Observing Network, aimed at measuring ice-sheet mass loss in western Antarctica.

Despite these problems, this IPY has yielded enough data for decades, and a solid baseline against which future observations can be compared. But Carlson says that the IPY must do more than just reveal the current state of the poles. The task ahead will be to translate observations into more reliable predictions.

"What we've seen during the IPY shows us

how unskilled our predictive ability still is," Carlson says. "We knew that the system is in change, but we're only beginning to realize how deeply."

Lost chances

The legacy of the IPY brings with it a few serious challenges. It remains unclear, for example, whether there will be enough support to create two proposed systems of polar monitors, the Sustaining Arctic Observatory Networks and the Southern Ocean Observing System, its Antarctic counterpart. Such basin-wide ground-observation networks are crucial for improving researchers' ability to model the ice caps and the heat flux that enters and leaves the oceans. Measurements of the thickness of sea ice, for example, are currently available from only seven points in the Arctic. And large parts of Antarctica's interior and surrounding oceans are even less well monitored.

Moreover, a functioning data archive, which would allow scientists around the world to access all data and observations collected during the IPY, is not yet in place. "In terms of data management we just didn't achieve what we wanted," says Carlson. "That's a big failure on our part for which I take responsibility."

The IPY data-management committee is sorting out the options for setting up a fully integrated data-sharing system. But the polar research communities hope that it will also be possible to maintain the general momentum and level of enthusiasm.

"Truth is there's not enough money to sustain everything," says Carlson. Therefore, he says, nations must invest in the most crucial measurements. Top on the wish lists of many scientists would be a permanent system to measure ice thickness in the Arctic.

Even as the IPY draws to a close, nations are already attempting to secure its legacy. On 22 February, environment ministers from Norway, Russia, Denmark, Sweden, China and Britain visited the Troll station to welcome back the participants of the Norwegian-led Antarctica traverse. The gathering, says Winther, served as a reminder to politicians that even after the end of the IPY, they must not forget the scientific value of the poles.

"There's a lot we could celebrate," he says. "But what's on our radar, much more than celebration, is continuity. Without a strong legacy, a lasting bonus, the beautiful lessons we've learned will be of much less value. We've seen the fruits of outreach and cooperation between research councils — we must not fall back into isolated efforts and national funding."

Just as important, says Carlson, will be to nourish the enthusiasm of the thousands of young scientists who participated in the IPY, and to secure the funding they need to pursue careers in research.

There is, indeed, no lack of interest and ambition. Since its foundation in 2006, the Association of Polar Early Career Scientists, a grass-roots organization of young polar researchers, has grown to some 1,400 members worldwide.

"We all have been hugely inspired by the IPY," says Jenny Baeseman, an adjunct assistant professor at the International Arctic Research Center in Fairbanks, Alaska, and co-founder of the association. "Now we want to stay involved in the science and in the political debate." ■

Quirin Schiermeier is *Nature*'s Germany correspondent.

> **"We all have been hugely inspired by the IPY. Now we want to stay involved in the science and in the political debate."**
> — Jenny Baeseman

The social pole?

As change in the Arctic accelerates, scientists and indigenous peoples have pressing reasons to work together, reports **Richard Monastersky**.

Indigenous peoples in the Arctic have long complained that the weather doesn't behave the way it used to. Climate scientists have by and large ignored them — until a few researchers looked into the data and found hints that the locals knew what was what.

"In the high latitudes, there have been a couple of studies in the past few years that find some support for the contention that the weather is becoming more variable, less predictable," says John Walsh, a professor of climate change at the University of Alaska at Fairbanks. In a 2005 study, Walsh and his colleagues uncovered evidence in the records of surface weather stations that Alaska and northern Canada experienced high-temperature extremes more frequently in the 1990s than they did from the 1950s to the 1980s (J. E. Walsh *et. al. Atmos. Ocean* **43,** 213–230; 2005). Such a change would indeed make the weather less predictable for the people who live there.

Walsh thinks that he and other scientists were too quick to write off the traditional knowledge of Arctic residents. Indigenous people have a deep understanding of their environment, and researchers must start paying more attention to what they say to catch the changes speeding through the far north, says Walsh. "We need to take a hard look [at local claims], and do the confirmation where it's feasible," he says.

Polar researchers of the past generally treated Arctic peoples as data points — or simply ignored them. Scientists would drop into a region, grab their measurements and vanish, rarely letting the residents know the results of the study. Now the relationship is getting more complex, for a raft of reasons. As well as seeing native peoples as potential collaborators and informants, scientists are also starting to see the importance of treating them as stakeholders with an interest in the results. And in some cases, governments led by indigenous people hold political power and must sign off research proposals.

In the hope of bridging the gap between scientists and the native inhabitants, countries participating in the Fourth International Polar Year (IPY), which finishes in March, supported some 30 projects that sought to tap into traditional knowledge of the Arctic. David Carlson, who heads the IPY programme office in Cambridge, UK, says that the IPY "set out with the idealistic goal of having the northern residents be partners in the science — not just subjects of study".

Such a partnership might build up research capacity among the indigenous population, advance scientific understanding of the Arctic and help native populations to adapt to climate change. Still, both scientists and native peoples say it is too soon to tell whether the IPY projects will bring much lasting

> **"Residents want to make their voices heard."**
> — Iver Campbell

Researchers think their studies could help mitigate the effects of climate change on herding societies.

change and will benefit both interest groups.

Victoria Gofman is optimistic on the subject. She is the executive director of the Aleut International Association based in Anchorage, Alaska, and principal investigator on the Bering Sea Sub-Network Project, a two-year IPY project surveying more than 600 residents in six villages bordering the Bering Sea — three in Russia and three in Alaska. The project members are asking local subsistence hunters about environmental and wildlife conditions in the Bering Sea and how they are changing. The aim, Gofman says, is to find a way that local knowledge can be integrated into data sets. The US National Science Foundation is funding the work to the tune of US$600,000.

Local benefits

The project is already starting to make a difference in the 650-person village of Gambell, on St Lawrence Island, Alaska. Iver Campbell, vice-chairman of the native village council, has interviewed about 50 residents for the project; he says they appreciated the chance to share their knowledge. "They want to make their voices heard. Everybody wants to be a part of this investigation of global warming." They are also curious to learn what scientists know about the changing conditions, he says.

Branson Tungiyan, the general manager for the native village, has a different take: he says the residents are tired of being studied and that the Bering Sea study will not significantly aid locals. "We don't see much benefit for Gambell," he says. How much benefit governments get, in terms of information that scientists and officials can use, remains to be seen, as the project is still gathering and analysing its data. But Gofman argues that the project has already helped the local communities, most obviously by employing people in the villages as data collectors at various rates of pay, and defends the idea of such work as a good vehicle for transferring traditional knowledge in a time of change.

In the Canadian territory of Nunavut, a study of polar bears is yielding more obvious mutual benefits for scientists and the native inhabitants, say participants. Moshi Kotierk, a co-investigator on the project, says the IPY has helped local communities by involving indigenous residents in the work, either as research assistants or, in his case, as a project leader. He thinks that this sort of approach will build the capacity for future scientific work in these areas.

Elizabeth Peacock, the Nunavut government's polar-bear biologist and principal investigator for the study, says she often relies on local residents and their knowledge of polar-bear behaviour. Her Inuit friends and contacts can tell her the locations of polar-bear dens, for example, or explain odd behaviour. Recently, she was puzzled when a male bear with a satellite transmitter stopped moving for six weeks, acting like a female in her den rather than hunting seals as would normally be the case. An Inuit hunter solved the mystery by telling her that male bears sometimes rest if they are already fat and want to preserve their energy for the best seal season later in the winter. That's not an insight you'll find in the scientific literature. "It took me talking to a hunter to understand that," Peacock says.

IPY leaders point to another successful

> **"It took me talking to a hunter to understand."**
> — Elizabeth Peacock

cross-cultural collaboration, studying reindeer herders and their vulnerability to change across Eurasia. The project, known as Ealát, after the word for 'good pasture' in the Saami language spoken in Lapland and beyond, is run by Ole Henrik Magga at Saami University College in Kautokeino, Norway, and Svein Mathiesen of the Norwegian School of Veterinary Science in Oslo. Some 3 million reindeer sustain hunters and herders in more than 20 different ethnic groups across northern Eurasia. The Ealát team is interviewing herders from a number of these groups, studying the climate at local scales, and investigating the political, social and economic factors that affect herders and their ability to adapt. By amassing such data, the researchers hope to mitigate the effects of climate change on herding societies.

Because herders are so central to the research, the Ealát project has spurred an interest in science within the herding community, says Johan Mathis Turi, secretary-general of the Association of World Reindeer Herders and a herder himself. "Young reindeer herders are pursuing higher education and are being included in knowledge production in a way we haven't seen in reindeer society before," he says.

Currently, five students from herding communities are pursuing doctoral degrees as part of the Ealát project. That will help herders interact with policy-makers, says Magga. "We need to strengthen the communication between herding communities and the government."

With data from projects still flowing in, scientists cannot say how broadly, if at all, they will incorporate local knowledge into their data sets and analyses, and how much improvement they may see. Outside the polar community, there are grumblings about including traditional knowledge in scientific studies. Two Canadian researchers published a book last year challenging government policies that require resource managers in the northern territories to take such indigenous knowledge into account (F. Widdowson and A. Howard *Disrobing the Aboriginal Industry* McGill-Queen's University Press, 2008).

That argument, however, does not win support within the polar-year leadership. Carlson applauds the projects that have involved native populations in scientific research, paving the way for future collaborations. "We would say our short-term success has been very high," he says, "and that the challenge remains to convert that into a long-term success."

Richard Monastersky is a features editor for *Nature* **in Washington DC.**

Scientists and native hunters are working together.

CORRESPONDENCE

Human dignity must be basis for debate on primate research

SIR — Bill Crum emphasizes a fundamental keynote of biomedical-research ethics in his Correspondence 'It should be possible to replace animals in research' (*Nature* **457,** 657; 2009) by stating that "good medical science" is not necessarily "morally justifiable or morally acceptable". On the other hand, many states and societies claim 'freedom of research' — meaning research being free from the need for justification — as a basic right. On the face of it, this looks like a discrepancy.

However, we have to recognize the fact that this freedom, like every other kind of freedom, has its ethical limits. Research can only be a right as long as it is not acting against our fundamental moral value: respect for human dignity. This is the basic point that we should agree on, regardless of our different opinions on what might constitute a breach of that principle.

With human dignity in mind, the ethical discussion about research on non-human primates has to focus on answering two questions. First, would prohibiting studies on primates constitute a threat to the human dignity of future generations, by reducing their chances of what we could consider a good life, as Roberto Caminiti states in his Correspondence 'Replacement of animals in research will never be possible' (*Nature* **457,** 147; 2009)? Second, is performing "invasive medical experiments" on creatures that "provide excellent experimental models of human cognition", as Crum states, a threat to our own dignity and our vision of how a good life should be led?

Only by using human dignity as the normative correlate for ethical decisions can we ensure that these decisions will be made on a basis that is equally important to all parties in this debate.

Tim Fieblinger Basal Ganglia Pathophysiology Unit, Lund University, BMC F11-46, 221 84 Lund, Sweden
e-mail: tim.fieblinger@med.lu.se

Readers are welcome to comment at http://tinyurl.com/c62pgf

Rare BSE mutation raises concerns over risks to public health

SIR — Atypical forms (known as H- and L-type) of bovine spongiform encephalopathy (BSE) have recently appeared in several European countries as well as in Japan, Canada and the United States. This raises the unwelcome possibility that variant Creutzfeldt–Jakob disease (vCJD) could increase in the human population.

Of the atypical BSE cases tested so far, a mutation in the prion protein gene (*PRNP*) has been detected in just one, a cow in Alabama with BSE; her healthy calf also carried the mutation (J. A. Richt and S. M. Hall *PLoS Pathog.* **4,** e1000156; 2008). This raises the possibility that the disease could occasionally be genetic in origin. Indeed, the report of the UK BSE Inquiry in 2000 suggested that the UK epidemic had most likely originated from such a mutation and argued against the scrapie-related assumption.

Such rare potential pathogenic *PRNP* mutations could occur in countries at present considered to be free of BSE, such as Australia and New Zealand. So it is important to maintain strict surveillance for BSE in cattle, with rigorous enforcement of the ruminant feed ban (many countries still feed ruminant proteins to pigs). Removal of specified risk material, such as brain and spinal cord, from cattle at slaughter prevents infected material from entering the human food chain.

Routine genetic screening of cattle for *PRNP* mutations, which is now available, could provide additional data on the risk to the public. Because the point mutation identified in the Alabama animals is identical to that responsible for the commonest type of familial (genetic) CJD in humans, it is possible that the resulting infective prion protein might cross the bovine–human species barrier more easily. Patients with vCJD continue to be identified. The fact that this is happening less often should not lead to relaxation of the controls necessary to prevent future outbreaks.

Malcolm A. Ferguson-Smith Cambridge University Department of Veterinary Medicine, Madingley Road, Cambridge CB3 0ES, UK
e-mail: maf12@cam.ac.uk
Jürgen A. Richt College of Veterinary Medicine, Kansas State University, K224B Mosier Hall, Manhattan, Kansas 66506-5601, USA

Scientific links with Cuba flourished despite US embargo

SIR — In your Editorial 'Cuba's biotech boom' (*Nature* **457,** 130; 2009), you state that "despite many constraints on interaction between Cuban and US scientists, biotech has prospered". In fact, US biotechnologists contributed in no small way to its development.

At the start, during the early 1980s, Cuban biotechnology was confined to a small house in a Havana suburb. An American group organized by Harlyn Halvorson, then director of Brandeis University's Rosenstiel Center and an inspirational leader, stepped in to help the venture. We were received warmly in Cuba whenever we visited.

The biotechnology effort soon transferred to a larger house across the street and from 1986 was housed in the majestic Center for Genetic Engineering and Biotechnology. The Cuban scientists set up symposia where one or more of us would speak.

The US government allowed us to travel to Cuba on the condition that we spent no American dollars there. We therefore continued to advise this fledgling group until the Soviet Union ceased to support Cuba financially and they could no longer pay for our visits.

Arnold L. Demain Research Institute for Scientists Emeriti, Drew University, Madison, New Jersey 07940, USA
e-mail: ademain@drew.edu

Idea of a love drug was no mystery to Shakespeare

SIR — In his Essay 'Love: neuroscience reveals all' (*Nature* **457,** 148; 2009), Larry Young claims that the biochemical understanding of love is not poetry. But at least one poet, namely William Shakespeare, foretold the application of drugs to manipulate the brain systems associated with pair bonding.

In *A Midsummer Night's Dream*, Oberon maintains that topical applications of the juice of the wild pansy (*Viola tricolor*, called 'love-in-idleness' in the play) "Will make or man or woman madly dote Upon the next live creature that it sees" (Act 2, Scene 1). The potion proves highly effective, supplying much of the humour in the play as Titania falls in love with the donkey-headed Bottom. Shakespeare also suggests that other substances from "Dian's bud" — variously identified as a species of wormwood (*Artemisia* spp.) or chaste tree (*Vitex agnus-castus*, a species not native to England but long known for its anti-libidinal properties) — could reverse the neurobiological results of the pansy. Perhaps poets have something to teach us about neurobiology and love after all.

Joan G. Ehrenfeld Department of Ecology, Evolution and Natural Resources, SEBS, 14 College Farm Road, New Brunswick, New Jersey 08901, USA
e-mail: ehrenfel@rci.rutgers.edu

Contributions may be submitted to correspondence@nature.com.

COMMENTARY

Man, machine and in between

Brain-implantable devices have a promising future. Key safety issues must be resolved, but the ethics of this new technology present few totally new challenges, says **Jens Clausen**.

We are so surrounded by gadgetry that it is sometimes hard to tell where devices end and people begin. From computers and scanners to multifarious mobile devices, an increasing number of humans spend much of their conscious lives interacting with the world through electronics, the only barrier between brain and machine being the senses — sight, sound and touch — through which humans and devices interface. But remove those senses from the equation, and electronic devices can become our eyes and ears and even our arms and legs, taking in the world around us and interacting with it through man-made software and hardware.

This is no future prediction; it is already happening. Brain–machine interfaces are clinically well established in restoring hearing perception through cochlear implants, for example. And patients with end-stage Parkinson's disease can be treated with deep brain stimulation (DBS) (see 'Human brain–machine applications'). Worldwide, more than 30,000 implants have reportedly been made to control the severe motor symptoms of this disease. Current experiments on neural prosthetics point to the enormous future potential of such devices, whether as retinal or brainstem implants for the blind or as brain-recording devices for controlling prostheses[1].

Non-invasive brain–machine interfaces based on electroencephalogram recordings have restored communication skills of patients 'locked in' by paralysis[2]. Animal research and some human studies[3] suggest that full control of artificial limbs in real time could further offer the paralysed an opportunity to grasp or even to stand and walk on brain-controlled, artificial legs, albeit likely through invasive means, with electrodes implanted directly in the brain.

Future advances in neurosciences together with miniaturization of microelectronic devices will make possible more widespread application of brain–machine interfaces. Melding brain and machine makes the latter an integral part of the individual. This could be seen to challenge our notions of personhood and moral agency. And the question will certainly loom that if functions can be restored for those in need, is it right to use these technologies to enhance the abilities of healthy individuals? It is essential that devices are safe to use and pose few risks to the individual. But the ethical problems that these technologies pose are not vastly different from those presented by existing therapies such as antidepressants. Although the technologies and situations that brain–machine interfacing devices present might seem new and unfamiliar, most of the ethical questions raised pose few new challenges.

Welcome to the machine

In brain-controlled prosthetic devices, signals from the brain are decoded by a computer that sits in the device. These signals are then used to predict what a user intends to do. Invariably, predictions will sometimes fail and this could lead to dangerous, or at the very least embarrassing, situations. Who is responsible for involuntary acts? Is it the fault of the computer or the user? Will a user need some kind of driver's licence and obligatory insurance to operate a prosthesis?

Fortunately, there are precedents for dealing with liability when biology and technology fail to work. Increasing knowledge of human genetics, for example, led to attempts to reject criminal responsibility that were based on the inappropriate belief that genes predetermine actions. These attempts failed, and neuroscientific pursuits seem similarly unlikely to overturn views on human free will and responsibility[4]. Moreover, humans are often in control of dangerous and unpredictable tools such as cars and guns. Brain–machine interfaces represent a highly sophisticated case of tool use, but they are still just that. In the eyes of the law, responsibility should not be much harder to disentangle.

But what if machines change the brain? Evidence from early brain-stimulation experiments done half a century ago suggests that sending a current into the brain may cause shifts in personality and alterations in behaviour. Many patients with Parkinson's disease who have motor complications that are no longer manageable through medication report significant benefits from DBS. Nevertheless, compared with the best drug therapy, DBS for Parkinson's disease has shown a greater incidence of serious adverse effects such as nervous system and psychiatric disorders[5] and a higher suicide rate[6]. Case studies revealed hypomania and personality changes of which the patients were unaware, and which disrupted family relationships before the stimulation parameters were readjusted[7].

Such examples illustrate the possible dramatic side effects of DBS, but subtler effects are also possible. Even without stimulation, mere recording devices such as brain-controlled motor prostheses may alter the patient's personality. Patients will need to be trained in generating the appropriate
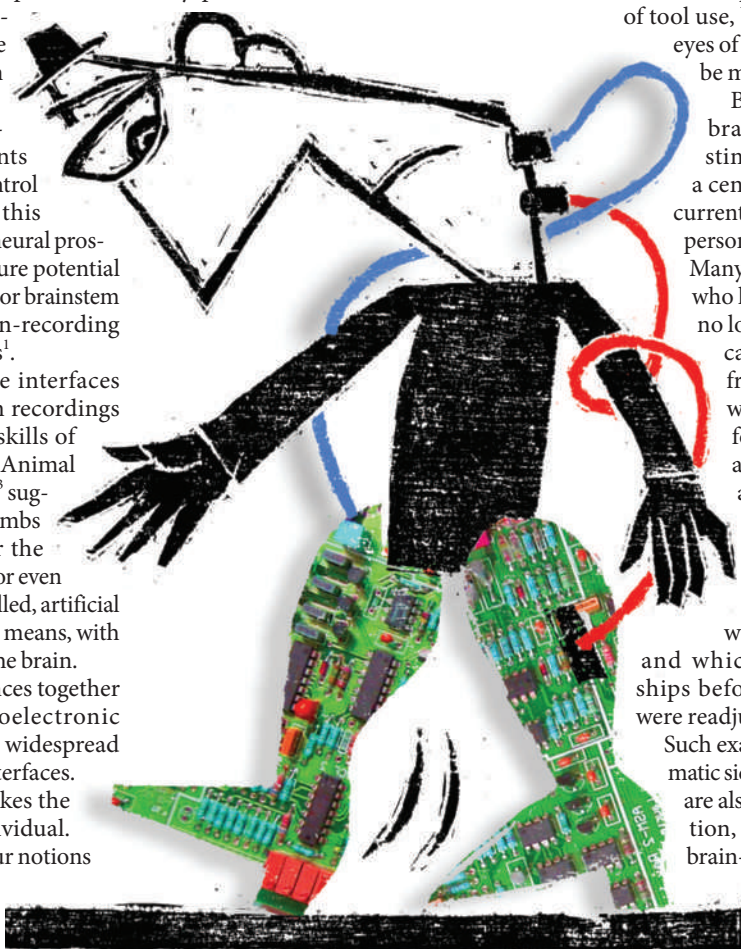
neural signals to direct the prosthetic limb. Doing so might have slight effects on mood or memory function or impair speech control.

Nevertheless, this does not illustrate a new ethical problem. Side effects are common in most medical interventions, including treatment with psychoactive drugs. In 2004, for example, the US Food and Drug Administration told drug manufacturers to print warnings on certain antidepressants about the short-term increased risk of suicide in adolescents using them, and required increased monitoring of young people as they started medication. In the case of neuroprostheses, such potential safety issues should be identified and dealt with as soon as possible. The classic approach of biomedical ethics is to weigh the benefits for the patient against the risk of the intervention and to respect the patient's autonomous decisions[8]. This should also hold for the proposed expansion of DBS to treat patients with psychiatric disorders[9].

Bench, bedside and brain

The availability of such technologies has already begun to cause friction. For example, many in the deaf community have rejected cochlear implants. Such individuals do not regard deafness as a disability that needs to be corrected, instead holding that it is a part of their life and their cultural identity. To them, cochlear implants are regarded as an enhancement beyond normal functioning.

What is enhancement and what is treatment depends on defining normality and disease, and this is notoriously difficult. For example, Christopher Boorse, a philosopher at the University of Delaware in Newark, defines disease as a statistical deviation from "species-typical functioning"[10]. As deafness is measurably different from the norm, it is thus considered disease. The definition is influential and has been used as a criterion for allocation of medical resources[11]. From this perspective, the intended medical application of cochlear implants seems ethically unproblematic. Nevertheless, Anita Silvers, a philosopher at San Francisco State University in California and a disability scholar and activist, has described such treatments as a "tyranny of the normal"[12], designed to adjust people who are deaf to a world designed by the hearing, ultimately implying the inferiority of deafness.

Although many have expressed excitement at the expanded development and testing of brain–machine interface devices to enhance otherwise deficient abilities, Silvers suspects that prostheses could be used for a "policy of normalizing". We should take these concerns seriously, but they should not prevent further

> "Electronic devices can become our eyes and ears and even our arms and legs."

Human brain–machine applications

Deep brain stimulation (DBS)
Electrical signals generated in a central computing unit (CCU) placed subcutaneously are sent to electrodes implanted deeply in the brain to stimulate specific structures, such as the subthalamic nucleus in patients with Parkinson's disease.
    Research directions: DBS is being extended to earlier stages of Parkinson's disease and to some other neurological impairments, including epilepsy, minimally conscious states and psychiatric disorders.

Cochlear implants
Acoustic signals are recorded by an external microphone and sent to a CCU placed outside the skull (containing a speech processor and a

radio transmitter) to generate electrical impulses. These are sent wirelessly to a receiver inside the skull and stimulate the auditory nerve at the internal neural interface — an electrode implanted into the cochlea in the inner ear.
    Research directions: for those whose auditory nerve is damaged, similar devices implanted into acoustically relevant areas of the brainstem or the midbrain are being tested clinically.

Retinal implants
Visual signals are recorded through light-sensitive diodes that stimulate the remaining cells of the retina. Alternatively, external signals recorded by camera are sent to an implant directly interfaced with the optic nerve. Some visual prostheses

bypass the retina and stimulate the visual cortex.
    Research directions: external power sources may be needed to provide the electrical stimulation required by some of these implants. Technologies for providing this power are being investigated.

Brain–machine interface-based motor prostheses
Motor signals are recorded through electrodes implanted in the brain and sent to a CCU, which computes and generates signals to direct an artificial limb.
    Research directions: animal experiments are becoming widespread. In experiments in patients with epilepsy, epicortical grid electrodes are placed on the surface of the brain to examine and improve decoding algorithms.

---

research on brain–machine interfaces.

Brain technologies should be presented as one option, but not the only solution, for paralysis or deafness. Still, whether brain-technological applications are a proper option remains dependent on technological developments and on addressing important safety issues.

One issue that is perhaps more pressing is how to ensure that risks are minimized during research. Animal experimentation will probably not address the full extent of psychological and neurological effects that implantable brain–machine interfaces could have. Research on human subjects will be needed, but testing neuronal motor prostheses in healthy people is ethically unjustifiable because of the risk of bleeding, swelling, inflammation and other, unknown, long-term effects.

People with paralysis, who might benefit most from this research, are also not the most appropriate research subjects. Because of the very limited medical possibilities and often severe disabilities, such individuals may be vulnerable to taking on undue risk. Most suitable for research into brain–machine interface devices are patients who already have an electrode implanted for other reasons, as is sometimes the case in presurgical diagnosis for epilepsy. Because they face the lowest additional risk of the research setting and will not rest their decision on false hopes, such patients should

be the first to be considered for research[13].

Brain–machine interfaces promise therapeutic benefit and should be pursued. Yes, the technologies pose ethical challenges, but these are conceptually similar to those that bioethicists have addressed for other realms of therapy. Ethics is well prepared to deal with the questions in parallel to and in cooperation with the neuroscientific research. ∎

Jens Clausen is at the Institute of Ethics and History in Medicine, University of Tübingen, Tübingen, Germany.
e-mail: jens.clausen@uni-tuebingen.de

1. Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S. & Schwartz, A. B. Nature **453,** 1098–1101 (2008).
2. Birbaumer, N. et al. Nature **398,** 297–298 (1999).
3   Hochberg, L. R. et al. Nature **442,** 164–171 (2006).
4. Greely, H. T. Minn. J. Law, Sci. Technol. **7,** 599–637 (2006).
5. Weaver, F. M. et al. J. Am. Med. Assoc. **301,** 63–73 (2009).
6. Appleby, B. S., Duggan, P. S., Regenberg, A. & Rabins, P. V. Mov. Disord. **22,** 1722–1728 (2007).
7. Mandat, T. S., Hurwitz, T. & Honey, C. R. Acta Neurochir. (Wien) **148,** 895–897 (2006).
8. Beauchamp, T. L. & Childress, J. F. Principles of Biomedical Ethics (Oxford Univ. Press, 2009).
9. Synofzik, M. & Schlaepfer, T. E. Biotechnol. J. **3,** 1511–1520 (2008).
10. Boorse, C. Phil. Sci. **44,** 542–573 (1977).
11. Daniels, N. Just Health Care (Cambridge Univ. Press, 1985).
12. Silvers, A. in Enhancing Human Traits: Ethical and Social Implications (ed. Parens, E.) 95–123 (Georgetown Univ. Press, 1998).
13. Clausen, J. Biotechnol. J. **3,** 1493–1501 (2008).

ESSAY

Microbial awakenings

A theory of how microbes 'wake up' from dormancy could help to solve scientific mysteries and improve disease control, says **Slava S. Epstein**.

Dormancy is a useful tool. Many plants and animals survive the winter in this way; spore-forming microbes use it to survive for perhaps millions of years. But what of non-sporulating species, which make up the great majority of microbial diversity?

Since at least the 1920s, these microbes have been strongly suspected to have a dormant state. But the picture isn't clear. Dormancy is exceedingly difficult to demonstrate directly, unequivocally prove or even to study in non-spore-forming microbial species. But why is it so hard to crack?

The answer may be simple: perhaps these microorganisms have a different mechanism of revival than those we are used to. Rather than awakening to environmental cues (as plants, animals and spores do), these microbes may wake stochastically — randomly.

I envision a situation in which microbial populations consist of a mix of active and dormant cells. Faced with an adverse environmental change, more cells are induced into dormancy, and survive the challenge. Individual cells would then periodically exit dormancy as a result of infrequent and essentially random events, such as a change in the expression of a master regulatory gene. I call such awakened cells 'scouts'. If the adverse conditions persist, the scout dies. If a scout forms under growth-permissive conditions, it starts a new population. In some species, scouts might even use growth-inducing signalling compounds to wake up the rest of the dormant population.

Waking the dead

Such an 'abnormal' pattern of revival, which differs from that of other microbes and contrasts with scientific expectations, would be easy to overlook. Researchers might see some evidence of revival events but, given the random nature of such events, be unable to reproduce the effect in further experiments, and so throw away legitimately positive results. Alternatively, researchers attempting to revive dormant microbes might instead find themselves accidentally and futilely attempting to wake up dead cells rather than dormant ones, thus providing a false negative.

The scout theory helps to explain apparently random disease recurrence, as happens for example with tuberculosis. Suppose that a small population of dormant cells escapes both the host's immune system and an antibiotic treatment (today's antibiotics target mainly actively growing cells). The possibility of disease recurrence would be determined by the balance between scout formation and the ability of the immune system to eliminate the scouts' progeny. When the pathogen succeeds, the relapse would seem to the observer as a random event.

If this is how microbes truly work, it has important implications. If the molecular mechanism of scout formation were known, researchers could potentially devise a way to make it less frequent, possibly to the point of rendering a population permanently dormant and thus harmless. Alternatively, they could artificially wake up all dormant cells into scouts, and dose them with drugs while they are active and vulnerable.

The scout model also redefines the nature of 'persisters', a small proportion of cells that are curiously tolerant of high concentrations of antibiotics and yet genetically identical to the vulnerable cells (by contrast with genetic mutants that might be resistant to a given antibiotic). Previous theories have proposed that microbial populations harbour specific, specialized dormant cells that survive the antibiotic challenge and grow once the challenge has been removed. It was unclear, however, how such persister cells could be both sufficiently dormant to not divide in an otherwise growing population the moment before the challenge, and then, under identical conditions after the challenge, be sufficiently active to form a colony.

My model postulates that the pool of dormant cells that survives the dose of antibiotic is substantially larger than the pool of persisters currently thought to survive. These dormant cells are effectively invisible to researchers because they do not grow, and so have been overlooked or dismissed as dead. If this pool of dormant cells continues to produce a small number of random scouts — before, during and after the antibiotic challenge — then the results are logical. Persisters are thus simply scouts generated by the large — and so far ignored — pool of true survivors.

Survival scouts

Dormancy and random scout formation could also help explain the 'great plate-count anomaly': the remarkable disparity between the high number of cells in a sample, and the small number of colonies these cells form *in vitro*. This was first observed in the late 1800s, making it arguably the oldest unresolved microbiological mystery.

Environmental samples harbour tens of thousands of species, most of which contain just a handful of cells per population. Low abundance suggests inactivity, and dormancy. Statistically, only a small minority of such rare and dormant populations would be expected to contain scouts at any given time. The majority of populations would therefore not grow, regardless of the conditions. In principle, scouts should appear in any population given a long enough period of incubation, but in practice Petri dishes are often thrown away by that time, or are overgrown by other, fast-growing species.

This suggests several ways to recover the 'missing' microbial diversity. Single cells could be cultivated, allowing them enough time for scout production unimpeded by faster growing species. Alternatively, researchers could attempt to identify signalling compounds used by scouts, if they exist, and use these molecules to induce growth of dormant kin.

The scout model makes other predictions. For example, in a prolonged incubation, a given microbial species would be expected to produce colonies not at a specific time but continuously over the entire duration of the growth experiment. These and other predictions can be checked through very simple experiments, several of which are under way in our laboratory. ∎

Slava S. Epstein is in the Department of Biology, Northeastern University, Boston, Massachusetts 02115, USA.
e-mail: s.epstein@neu.edu

D. PARKINS

BOOKS & ARTS

In Retrospect: the physics of sand dunes

Ralph Bagnold's wartime posting to map the dunes of North Africa was like a desert epic — and it inspired his classic text on how wind-blown grains self-organize into regular patterns, explains **Philip Ball**.

Physics of Blown Sand and Desert Dunes
by Ralph Bagnold
Methuen: 1941. 265 pp.

Ralph Bagnold intuited that regular dune and ripple shapes embody the physics of complex systems.

"We forgave Bagnold everything for the way he wrote about dunes. 'The grooves and the corrugated sand resemble the hollow of the roof of a dog's mouth.' That was the real Bagnold, a man who would put his inquiring hand into the jaws of a dog."

This is the voice of Hungarian explorer László Almásy, as fictionalized in Michael Ondaatje's 1992 novel *The English Patient*. The remark is a piece of authorial indulgence, for Ondaatje's readers could hardly be expected to know of this Bagnold who flits across a few pages of the desert epic. Ondaatje, however, clearly knew his man, hinting at the groundbreaking way in which Ralph Alger Bagnold studied and wrote about the major obsession of his life: desert sands.

Bagnold did so with such perception and insight that his 1941 book *The Physics of Blown Sand and Desert Dunes* became the standard work on dune formation for many decades. Informed by observations in Libya and wind-tunnel experiments in the United Kingdom, Bagnold set out to explain how sand grains are organized by wind into structures ranging from ripples the width of a finger to undulations several kilometres across.

His book is more than an exploration of a hitherto-neglected aspect of geomorphology. In hindsight, Bagnold's work can be seen as a landmark in a much broader vista: the investigation of complex, self-organizing systems, in which order and regularity emerge from interactions of components that seem to prescribe nothing more than chaos and disorder. As Bagnold wrote of dunes:

"The observer never fails to be amazed at a simplicity of form, an exactitude of repetition and a geometric order unknown in nature on a scale larger than that of crystalline structure."

Thus he perceives that the problem is not simply to explain the size, shape and variety of dune-like formations, but to account for the more general phenomenon of regularity arising from a system that is initially featureless.

Sand ripples and dunes are conspiracies of

"The observer never fails to be amazed at a simplicity of form, an exactitude of repetition."

grains that arise from the interplay of wind-borne transport, collision-driven piling up, and slope-shaving avalanches. They are an archetype for the self-organized patterning of systems of many interacting components. Interest in granular complexity has blossomed over the past decade, exemplified by the appearance of stationary-wave arrays in shaken, shallow layers; grain-size stratification in avalanches and rotating drums; and the use of sand piles as a model of 'scale-free' dynamics known as self-organized criticality. Such behaviour lies squarely in the domain of the physicist; Bagnold's prescience is located in his intuition that wind-formed geomorphology embodies the physics of complex systems writ large in nature.

There is a long tradition of army and naval engineers whose scientific legacy implies that their minds were on matters other than the military. Bagnold joined the British Army's Royal Engineers in 1915 and fell in love with the deserts during postings to Egypt and India. By the 1920s he was spending his leave exploring these 'seas of sand', joining Almásy's 1929 expedition in search of the legendary city of Zerzura west of the Nile.

Before Bagnold, the transport of small particles by fluids — not just sand in wind, but sand and silt in water, wind-blown snow and industrial processing of grains such as cereal and coal dust — was afforded little more than a few empirical formulae used by engineers.

But Bagnold claimed that "the subject of sand movement lies far more in the realm of physics than of geomorphology". He began from an aerodynamic perspective, conducting wind-tunnel studies in the mid-1930s to map the trajectories of individual sand grains in moving air. The key to the transport process is that the grains bounce when they hit the desert floor, and are carried along in a series of little jumps — known as saltation — that ultimately determine the length of sand ripples.

But the central issue was why wind-borne sand produces roughly regular ridges. Bagnold showed that this was caused by a self-amplifying growth instability, now recognized to drive patterning in systems such as snowflakes and dendrites, fractal aggregation and viscous fingering. He showed that a single, chance irregularity on a smooth desert floor stimulates growth and multiplication of bumps of more or less equal size and spacing.

He began fieldwork in Libya in 1938, funded by Britain's Royal Society. His book was published three years later, while he was conducting reconnaissance in the North African conflict during the Second World War. He deduced the wind conditions that produced different dune types, such as crescent-shaped barchans and undulating seif dunes — conclusions that have been recently borne out by computer modelling.

The general problem that Bagnold faced — to account for spontaneous pattern formation — was addressed more famously in D'Arcy Wentworth Thompson's seminal 1917 book

On Growth and Form. But Bagnold seemed unaware of that. Thompson's epic revised edition, published in 1942, surprisingly neglects not only Bagnold's efforts but the entire issue of ripple and dune formation. A connection to more general patterning processes ultimately emerged from Alan Turing's work on biochemical morphogenesis, described in a 1952 paper. When, in the 1970s, mathematical biologists Hans Meinhardt and Alfred Gierer identified the fundamental ingredients of Turing's stationary chemical patterns — the presence of a locally acting autocatalytic 'activator' and an inhibitor that suppresses pattern elements over longer

ranges — it became apparent why sand ripples resembling in plan form the striped markings of zebras probably result from a Turing-like mechanism. The formation of a ripple is self-enhancing because it captures more sand the bigger it gets. Meanwhile, this process depletes the air of sand grains, suppressing another ripple for some distance downwind.

The fact that granular flow might serve as a universal analogy for other physical phenomena had been suspected in the late nineteenth century by Osborne Reynolds, a pioneer of fluid dynamics. In order to flow, a collection of grains must expand a little, and Reynolds

decided that this 'dilatancy' of powders could explain all the mechanical behaviours in nature if space were filled with submicroscopic grains. A portrait from 1904 shows Reynolds holding a basin of ball bearings, and two years earlier he revealed what he had in mind: "I have in my hand the first experimental model Universe, a soft India rubber bag filled with small shot." William Blake's world in a grain of sand is invoked to the point of cliché in granular research, but here it was claimed as a reality. ■

Philip Ball is a consultant editor for *Nature*. His forthcoming book series is *Nature's Patterns: A Tapestry in Three Parts.*

Inside the map-maker's mind

I trace with my finger the ridgeline to the summit of Mount Everest. The beautiful, icy, white-, blue- and granite-coloured map on the cover of *The Natures of Maps* brings the peak easily within reach. Yet if I were to try to scale this mountain, it is likely I could die trying. In this sense, argue geographers Denis Wood and John Fels, this map puts nature in its place: under my thumb.

Although I know it is a representation of nature, and not the real thing, such representations are powerful. They affect how we think about the subjects they portray. And therein lies the utility of this terrific book. It uses the tools of cognitive linguistics to conduct a step-by-step analysis of how maps construct — in our minds — the versions of nature that dominate public discourse about the environment, ecology, conservation and the proper place of humans on our planet.

The authors identify eight versions of nature that are constructed by the arguments commonly embodied in maps. Nature may be awesome, a threat or a victim. It embodies a cornucopia, is collectable and an object of scientific study, yet it remains a mystery. Or it may be differentiated as a park, legally protected and codified, "a nature, ultimately, quietly put in its place".

The book is a beautiful tour de force. Laid out like an art book with stunning reproductions of maps, it also contains a trenchant, practical analysis that is useful for anyone wanting to read maps more

critically and construct better maps of nature.

Wood and Fels borrow their conceptual scaffold from cognitive scientists Gilles Fauconnier and Mark Turner, who argue that language opens 'mental spaces' that can blend with other conceptual spaces in our minds to create new combinations. For example, the term computer virus is a powerful mix of two disparate ideas: one technological, the other biological. Wood and Fels analyse how a map similarly provides "a system of propositions" about nature that "get tied together into arguments about the world".

These spatial arguments are constructed

The peak of Mount Everest, within touching distance.

'on the fly' in our minds, say Wood and Fels, presumably using the same kind of activated neuronal assemblies that are proposed to enable the conceptual blending of their linguistic counterparts. But this neuroscience black box distracts from their analysis of how "maps hoist themselves off the page into our brains, spawning world views" as we read the complex propositions posted on their flat surfaces.

The analyses of the eight natures commonly constructed by maps provide the book's greatest value. 'Threatened nature' is the most compelling, and Wood and Fels bring all their tools to bear in an incisive deconstruction of a map from *National Geographic* entitled 'Australia under siege'. They trace the argument being made in geographic terms as this standard, seemingly objective base map is blended — in the reader's mind — with colourful maps of Australia's land cover 200 years ago and today, showing the threats posed by fire, feral species, forestry, grazing and mining. Ultimately, they say, the map argues that the past equals nature, and no map of the future is needed: "the meaning (and the fear and anxiety) emergent in the blend is perfectly clear".

The eight natures arguably encompass the most important currents in contemporary thought, save for one: nature is change. That is not just the nature that has been changed, as in Australia, but nature that is always dynamic. The omission of this dynamism is a weakness of this book, and in fairness, of most maps. It is a pity that the authors limit their analysis to static maps of nature when we are witnessing the proliferation of 'mash-ups' that link data sources to web-based applications, such as Google Earth, to create dynamic, interactive maps. Fortunately, the analytical tools that Wood and Fels demonstrate can help us understand how interactive maps work too.

Dynamic maps open up new mental spaces more quickly and readily than static

maps, and can be generated by users who ask their own questions of data sources. Instead of going to pre-existing maps for answers about nature, we can create our own maps to query nature. Rather than relying on reified versions of idealized natures, maps can allow us to explore changing versions of the real, messy natures we live in.

The Natures of Maps should be read and put to use by anyone who makes or uses maps, whether they are scientists, conservationists or landowners. The constructs of nature that Wood and Fels identify inform the way maps are interpreted. And the analytical techniques that they deploy can be used to make new maps

of science and nature that are better at helping us to ask important questions.

Maps are indeed arguments about our world, but the future also rides on maps. People use them to shape what we know and what becomes of the territory. As Wood and Fels argue so provocatively, "Pretending to be no more than scorekeepers, maps stand revealed as more like the ball, the very medium through which the game's moves are made." ∎

Jon Christensen is associate director of the Spatial History Project at the Bill Lane Center for the American West, Stanford University, Stanford, California 94305-4225, USA.
e-mail: jonchristensen@stanford.edu

Who's watching you?

The Privacy Advocates: Resisting the Spread of Surveillance
by Colin J. Bennett
MIT Press: 2008. 259 pp. £18.95

Privacy: A Manifesto
by Wolfgang Sofsky
(Translated by Steven Rendall)
Princeton University Press: 2008.
148 pp. £13.95

If you have ever caught a bus, you will be familiar with this experience: you wait ages for one to arrive, then two turn up together. As with transport, so with books. These two titles use different approaches to unpick the entwined concepts of surveillance, privacy and human rights.

In *The Privacy Advocates*, political scientist Colin J. Bennett reviews the privacy landscape from the perspective of those who seek to prevent society from sliding into an unthinking acceptance of surveillance culture. Advocates of surveillance cameras in public places cite the prevention of crime as a key motivation, yet the same technology could be subverted to intrusively monitor innocent people.

Bennett suggests that we should all be privacy advocates because there are elements of everyone's life that are so personal that they should not become public or state property. Active himself in issues of privacy, Bennett could risk bias in covering such an emotive topic. Instead, he seeks to hold up a mirror to the network of privacy advocates. By allowing the network members he already knows to select secondary contacts for him, he develops an organic web of opinion and insight.

The strength of character of many privacy campaigners might set them as outliers of current social opinion. To his credit, Bennett

Banksy's graffiti art criticizes surveillance culture.

takes great care to avoid depicting them as a naive fringe element. In describing the publicity-friendly techniques used by the privacy movement, such as performing Shakespearean drama in front of surveillance cameras, he acknowledges the serious intent of such activity as a wake-up call to society.

In a thoughtful and determined voice, Bennett provides an international perspective and comprehensive coverage. The material is current and the history of the topic is well presented. The development of the privacy lobby is tied in with the technological advances that, to a large extent, have driven it — such as the

ease with which the origin of any Internet communication can be traced. It is daunting to see how quickly we have become inured to surveillance as normal when we would once have viewed it as intrusive. My daily bus journey is now almost entirely recorded on video — a development that was inconceivable both technically and ethically only a few years ago.

Privacy, a slim but beautifully produced volume by sociologist Wolfgang Sofsky, is a manifesto in the classic sense. In a 148-page translation from the original German edition of 2007, he outlines his concerns over the depth of information that is now held, analysed and used from almost every aspect of our lives. Beginning with a review of the daily surveillance encountered by an average citizen, the author takes us on a personal journey that discusses the cultural roots of privacy, the origins of property and the pivotal nature of freedom of thought.

Sofsky covers an enormous amount of territory on his voyage, and digs deep into our core social values to discuss the origins of our behaviours, interactions and innate needs. Human territoriality, the concepts of personal space and the conventions regarding bodily functions all build together into a convincing justification of why privacy becomes such an emotive issue in a crowded, and perhaps threatening, environment. Private spaces are our refuges from society at large. Sofsky discusses the way we extend private space — into our cars, for example — but also shows us how vulnerable we become to these boundaries being breached.

Setting this against the development of an information-oriented world, Sofsky skilfully weaves his argument that we, as groups and individuals, have lost significant independence and freedom of action. We are, he argues, at the behest of "an unholy alliance of institutions" — both governmental and commercial — that ensure we never feel free from observation in any aspect of our lives. The technologies are available, they are being used and, as the tools become more sophisticated, so too will the use of the data they generate.

Reading Bennett's book is like joining a conversation in a commuter train where the regular passengers know and respect each other. You are informed of the points being debated, then left to make your own judgement. The journey with Sofsky is similarly amicable but is more like being a passenger on a long taxi ride, where the driver shares his life-vision and philosophy. Both approaches work well — the time spent in the authors' company passes both quickly and profitably. ∎

John Gilbey is an honorary lecturer in the Department of Computer Science at Aberystwyth University, Aberystwyth SY23 3DB, UK.
e-mail: gilbey@bcs.org.uk

C. GILLON/GETTY IMAGES

Movies for a scientific mind

It will be a good year for films about science, judging from the screenings at the 2009 Sundance Film Festival, held in Park City, Utah, last month. Aside from environmental documentaries, some of the more intriguing films on offer examined the human mind.

The documentary *Boy Interrupted* chronicled the life of a teen with bipolar disorder who jumped to his death after his parents took him off lithium, and *Over the Hills and Far Away* followed a couple scouring Mongolia for a shaman to cure their son's autism. Two fiction films, starring Kevin Spacey and Chazz Palminteri, told the stories of psychiatrists negotiating their own mental breakdowns.

Some of the best films got inside the twisted minds of fighters. *Bronson* made theatre of the psychopathic exploits of a British prisoner, and the documentary *Tyson* wove the musings of the boxer into a portrait of an exquisitely vulnerable man. Not all films reached this standard. *Brief Interviews with Hideous Men*, about a female anthropologist studying how men view sex, did not retain the cranky charm of David Foster Wallace's novel. *Adam*, in which a schoolteacher falls for a young engineer with Asperger's syndrome, was stiff and preachy. Yet it won the Alfred P. Sloan prize for films depicting scientists.

"For a screenwriter it's always so much easier to tell a story about the perils of science than about incremental progress," said Jeffrey Nachmanoff, who co-wrote the 2004 global-warming thriller *The Day After Tomorrow* and who served on the Sloan prize jury. The festival was remarkably free of such sensationalism.

An unusually strong presence of science fiction included a coincidental pairing of movies about astronauts encountering their own cloned replacements. Although the Japanese *The Clone Returns Home* was rather slow, its British counterpart, *Moon*, directed by David Bowie's son Duncan Jones and starring Sam Rockwell and Kevin Spacey, was more entertaining. If one accepts their premise — that doctors will eventually be able to duplicate not just bodies but minds — these films raise questions about medical ethics and the origins of identity.

One astrophysicist expressed frustration that none of the 118 films at Sundance depicted an ordinary scientist at work, but not all agree. "I don't make such a distinction between pure science and science fiction," replied John Underkoffler, science adviser on the films *Minority Report* and *Iron Man*. "At their best films convey ideas, and the guise isn't so important." ■

Jascha Hoffman is a writer based in New York.

Q&A: Getting under Darwin's skin

Charles Darwin's relationship with his wife Emma is the subject of the film *Creation*. Actor **Paul Bettany** describes portraying Darwin's turmoil over his daughter's death and religious conflict with his wife.

Darwin**200**

Was it difficult playing such a famous man?
It was a real honour and horribly frightening in equal measure. Often when I prepare I must scratch around for research, but with Darwin there is almost too much material. He was such a prolific writer and there has been so much written about him that no matter how much I read, I felt that I hadn't done enough.

Paul Bettany as Charles Darwin in *Creation*.

L. DANIEL/ICON FILM DISTRIBUTION

Did your experience of playing the ship's surgeon in the 2003 film *Master and Commander* help you?
I found it really useful having spent time in the Galapagos Islands making that earlier movie. I was able to draw on a real place in my memory. *Creation* is set during the writing of *On the Origin of Species* 20 years after Darwin returned from the Galapagos. The film is about his procrastination and the reasons for it — his inability to write, it making him sick, the death of his daughter, the effect that had on his relationship with his wife, and their different means of dealing with it.

Scientists are often portrayed as being emotionless, so how did you draw out his intense grief following his daughter Annie's death?
People say he looked at his children like they were projects. But I think that this was a sign of the utmost love and respect. Science was his life. So of course he was going to involve his family. For him, science was an act of love, and study was an act of love. I don't think that science is a dry, dusty thing for scientists.

Was it easy playing opposite your wife, Jennifer Connelly, as Emma Darwin?
We got a lot for free in terms of the physicality that real married people have with each other. When you are trying to produce a relationship on screen that doesn't actually exist between the actors, there is a temptation to look at or touch each other more, whereas there is an absent-minded closeness between Jennifer and I in the movie.

How did you handle the conflict between Darwin's agnosticism and his wife's strong Christian beliefs?
Darwin was a social conservative who had a revolutionary idea, and it was very difficult for him. He couldn't help seeing how the idea fitted like a glove everywhere he looked, in the indifferent cruelty of nature. When their child died, he went to science and she went to religion. And the thing that he was working on was potentially going to take her solace away. So it is incredibly moving and dynamic to play.

How did they convey their feelings?
Charles and Emma had a fantastic system of writing. They were in the same house but would write each other letters to discuss things. They were brutally honest with each other about their feelings. Even so, Annie's death was very difficult for them to discuss.

Any plans to play him again?
I would love to play the old Darwin, simply to be able to say I've had the biggest beard in cinematic history. ■
Interview by **Adam Rutherford**, *Nature*'s podcast and video editor.

Listen to the full interview on the *Nature* podcast at www.nature.com/nature/podcast/index.html.

1087

NEWS & VIEWS

End of agriculture — an abandoned collective farm in Belarus.

GLOBAL CHANGE

Carbon in idle croplands

Geoffrey M. Henebry

The collapse of the Soviet Union had diverse consequences, not least the abandonment of crop cultivation in many areas. One result has been the vast accumulation of soil organic carbon in the areas affected.

The formal dissolution of the Soviet Union in December 1991 triggered the most widespread and abrupt episode of land change in the twentieth century. The institutional changes produced socio-economic dislocation throughout the former Soviet Union and its client states. The agricultural sector was hit particularly hard, with large tracts of cultivated land being abandoned and reverting gradually to grassland. This, in turn, had serious effects on the dynamics of carbon, water and biomass.

Writing in *Global Biogeochemical Cycles*, Vuichard and colleagues[1] describe a modelling study aimed at exploring the plausible trajectories of carbon accumulation, above and below ground, following land abandonment. Their study area spanned roughly 171,000 square kilometres in the agricultural regions of Ukraine, Belarus and the European portion of the Russian Federation.

The balance sheet for carbon accounting includes, first, the current stocks of carbon in a landscape, including live and dead plant parts above and below ground, an array of soil-dwelling organisms, and partially decomposed and transformed organic materials. The second component is the flux of carbon into and out of the landscape. Primary input of carbon comes through photosynthesis and the fixing of carbon dioxide into simple sugars — a process called primary production. Output includes CO_2 waste from the metabolic processes (respiration) of animals, plants and microbes as they go about the business of living in that landscape.

The net balance — primary production less respiration — constitutes the net ecosystem productivity (NEP). But there are other sources of flux to consider: the physical transport of organic materials by wind, water, humans and other animals, and the loss of carbon from the landscape through fire. Of these, it is the human use of the products of photosynthesis — food, fibre and wood — that results in transport of organic matter from the landscape in which it grew. It is the NEP of an area less its harvest (and eventual consumption elsewhere) that constitutes the net biome productivity (NBP) of a landscape or a region. In croplands the NBP is found mostly below ground, in the form of carbon that is locked into the soil as partially decomposed molecules that resist further transformation.

Vuichard and colleagues[1] linked models of ecosystem processes and of crop production to calculate the amount of carbon sequestered in the soil as grasses replaced crops. The region they targeted — 40° to 60° N by 20° to 60° E — has been estimated[2] to include much of the former Soviet Union's croplands abandoned during the 1990s. To calculate the NBP resulting from cessation of cultivation, the authors confronted several challenges. I will touch on only two.

First, the models required estimates for growing-season weather at high temporal resolution. Vuichard *et al.* accomplished this by adding six-hourly variability data to monthly averages and then re-sampling to half-hour resolution. Although this recipe can produce a credible reconstitution, the weather is often more quixotic than our statistical cartoons permit. Second, simulating cultivation regimes during the Soviet period required detail about the management practices applied. The authors finessed a dearth of data by extrapolating trends in fertilizer applications, estimating irrigation applications, and delimiting the possible outcomes with three management extremes: no added nitrogen, no tillage, or limited crop-residue inputs to the soil.

Running the coupled models to simulate the abandoned croplands from 1991 to 2000, Vuichard and colleagues estimate that carbon was sequestered in soil on a grand scale. The net accumulation rate, however, has fallen from an initial annual high of 105 grams of carbon per square metre to a mean of 47 ± 33 g C m^{-2} yr^{-1}. These results compare favourably with rates from an empirical study[3] of a small number of fields at an experimental station near Moscow: 132 ± 21 g C m^{-2} yr^{-1} for the first 15 years after the cessation of cultivation, decreasing thereafter to 70 ± 8 g C m^{-2} yr^{-1}. The spatial pattern of NBP showed the greatest accumulation in the northwestern sector of the study region, decreasing sharply to the southeastern sector, where dry conditions led to net carbon loss rather than gain.

Overall, Vuichard and colleagues[1] calculate a carbon sink in the abandoned croplands of the former Soviet Union of 8 Tg C yr^{-1}, which pales in comparison with the sink of roughly 270 Tg C yr^{-1} for the vast Russian forests[4]. However, the sink strength of the abandoned land (47 g C m^{-2} yr^{-1}) is about 50% greater than that
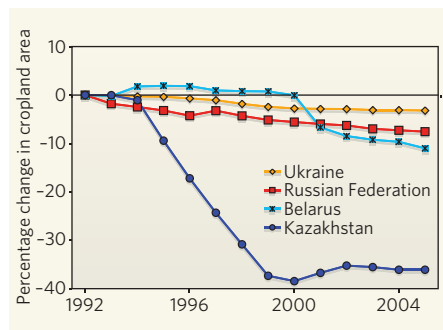
Figure 1 | Change of the area in arable land and permanent crops following the collapse of the Soviet Union. The relative abandonment in Kazakhstan is much greater than in the Russian Federation, but the areal difference is only 25% greater (127,010 km² compared with 101,270 km²). Note the initial increase in croplands in Belarus following independence and the sharp contraction after 2000. (Baseline year is 1992. Source ref. 15.)

of the Russian forests (31 g C m⁻² yr⁻¹) (ref. 5). In contrast, the average annual carbon sink for all of Europe between 1990 and 2100 has been projected at 17–38 Tg C yr⁻¹, or merely 2–3% of the European Union's CO_2 emissions over the same period[5].

The magnitude of the estimated accumulation is sensitive to the cropland management practices before the abandonment of agriculture. But, in the absence of better ground data, the modellers must resort to a 'sensitivity analysis' that explores possible rather than likely outcomes. Given the current interest in Europe and North America in turning annual croplands into perennial grasslands for the production of cellulosic ethanol, or to sequester carbon, this study offers a cautionary note: land-use history must be considered when framing and refining potential carbon sinks.

Vuichard and colleagues' focus on European Russia misses a more complicated story of agricultural land-use change in the wake of the break-up of the Soviet Union[6,7]. Cessation of agriculture has been patchy in Russia, resulting in an archipelago of communities that are threatened by isolation[8]. Moreover, abandonment of arable land was far more prevalent in Kazakhstan[9], the heart of the Virgin Lands Program launched in the mid-1950s by the Soviet leader Nikita Khrushchev (Fig. 1). This led to another profound period of land-cover change aimed at reducing the risk of famine by spreading cereal production into Asian Russia[10].

Finally, this study[1] underscores the need to improve representations of land-cover and land-use dynamics if we expect to provide credible guidance about effective strategies for adapting to and mitigating climate change[11]. The incorporation of agriculture into climate models continues to advance[12,13], but considerable challenges remain. To capture the spatial heterogeneity and temporal dynamics of human land-use, we need to move beyond the static land-cover classes that were once a solution to the problem of limited observational data. Viewed from orbital sensors,

the growth and development of vegetation across the land surface reveals distinctive patterns — 'land-surface phenologies' — that relate to land use, vegetation type, climate and recent weather, and can indicate land-cover change[9]. Assimilating land-surface phenologies into models can forge the link that is required to feed back the changing state of vegetation to atmospheric processes[14]. ■

Geoffrey M. Henebry is at the Geographic Information Science Center of Excellence, South Dakota State University, Brookings, South Dakota 57007-3510, USA.
e-mail: geoffrey.henebry@sdstate.edu

1. Vuichard, N., Ciais, P., Belelli, L., Smith, P. & Valentini, R. Glob. Biogeochem. Cycles doi:10.1029/2008GB003212 (2008).
2. Hurtt, G. et al. Glob. Change Biol. 12, 1208–1229 (2006).
3. Kurganova, I. N. et al. Žemés Ükio Mokslai. 15, 10–15 (2008).
4. Shvidenko, A. & Nilsson S. Tellus B 55, 391–415 (2003).
5. Zaehle, S. et al. Ecosystems 10, 380–401 (2007).
6. Lerman, Z., Csaki, C. & Feder, G. Agriculture in Transition: Land Policies and Evolving Farm Structures in Post-Soviet Countries (Lexington Books, 2004).
7. Kuemmerle, T., Hostert, P., Radeloff, V. C., Perzanowski, K. & Kruhlov, I. Ecosystems 11, 614–628 (2008).
8. Ioffe, G., Nefedova, T. & Zaslavsky, I. Ann. Assoc. Am. Geogr. 94, 913–943 (2004).
9. de Beurs, K. M. & Henebry, G. M. Remote Sens. Environ. 89, 497–509 (2004).
10. McCauley, M. Khrushchev and the Development of Soviet Agriculture, The Virgin Land Programme 1953–1964 (Bowering, 1976).
11. Betts, R. Tellus B 59, 602–615 (2007).
12. Osborne, T. M. et al. Glob. Change Biol. 13, 169–183 (2007).
13. Bondeau, A. et al. Glob. Change Biol. 13, 679–706 (2007).
14. Adegoke, J. O., Pielke, R. Sr & Carleton, A. M. Agric. Forest Meteorol. 142, 203–215 (2007).
15. UNFAO. ResourceSTAT-Land (December 2007) http://faostat.fao.org/site/377/default.aspx#ancor.

ALZHEIMER'S DISEASE

A prion protein connection

Moustapha Cisse and Lennart Mucke

More than 20 million people worldwide have Alzheimer's disease, yet its causes remain mostly uncertain. Fresh findings provide molecular clues, linking this disease to another neurodegenerative disorder.

Investigations of the causes of Alzheimer's disease yield one culprit time and time again: abnormal build-up of amyloid-β (Aβ) peptides in the brain. Small, soluble aggregates of Aβ — Aβ oligomers — impair memory by disrupting memory-related functions of synaptic junctions between neurons[1–3]. But whether specific receptors mediate these adverse effects has remained unknown. In this issue, Laurén et al.[4] (page 1128) show that the prion protein might mediate the pathogenic effects of Aβ oligomers.

The prion protein (PrP) is anchored to the cell membrane and associates with membrane microdomains called lipid rafts. It occurs in at least two conformational states. The cellular form, PrP^C, is involved in maintaining the brain's white matter, and in regulating this tissue's innate immune cells, responses to oxidative stress and neuron formation[5]. The highly pathogenic form, PrP^Sc, is a misfolded version of PrP^C and is resistant to enzymatic degradation. PrP^Sc is the main cause of a group of fatal neurodegenerative disorders called transmissible spongiform encephalopathies that includes Creutzfeldt–Jakob disease and mad cow disease[6].

Aβ binds to and influences the function of many cellular proteins[7]. Laurén et al.[4] therefore set out to identify proteins with greater affinity for Aβ oligomers than for freshly solubilized, putatively monomeric and non-toxic Aβ. Their unbiased genome-wide screen yielded PrP. Interaction between Aβ oligomers and PrP did not require the PrP^Sc conformation, although

the authors did not explore whether PrP^C misfolds into a PrP^Sc-like conformation on binding to Aβ oligomers.

Synaptic connections between neurons can be strengthened through a phenomenon called long-term potentiation (LTP), which provides a measure of synaptic plasticity related to learning and memory — two faculties that are compromised in Alzheimer's disease. To investigate what effect Aβ–PrP^C interaction might have on LTP, Laurén et al. studied this process in slices of mouse hippocampus, a brain region crucial to learning and memory. They find that Aβ oligomers inhibit LTP in hippocampal slices from normal mice, but not in hippocampal slices from mice lacking PrP^C. Similarly, LTP was not affected by Aβ oligomers in hippocampal slices from normal mice in which Aβ–PrP^C interaction was blocked. So PrP^C seems to be a main receptor for Aβ oligomers, mediating their deleterious effects on synaptic function.

The authors appropriately note, however, that they cannot exclude the existence of other receptors for Aβ oligomers, because PrP ablation reduced the binding of Aβ oligomers to neurons by only 50%. Alternative receptors might include the transmembrane proteins APLP1 and 30B, but, compared with PrP^C, both of these showed much lower affinity and selectivity for Aβ oligomers[4]. Similarly, compared with PrP^C, another Aβ-binding protein, RAGE, showed much lower affinity and selectivity for Aβ oligomers. However, earlier work indicated that disrupting Aβ–RAGE interaction inhibits programmed cell death in
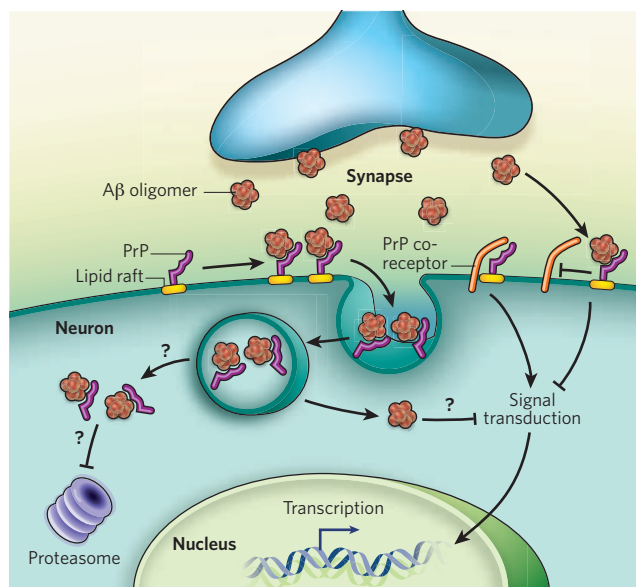
Figure 1 | PrP and Aβ-induced synaptic toxicity. Laurén *et al.*[4] show that Aβ oligomers interact with the membrane-bound prion protein PrP^C. This interaction may in turn disrupt interaction between PrP^C and a co-receptor, impairing the neuron's signal-transduction pathways required for synaptic plasticity. Alternatively, internalization of PrP^C may allow Aβ oligomers to reach intracellular compartments, where they might interfere with cellular functions such as protein degradation by the proteasome complex, and PrP^C-dependent gene transcription.

cortical neurons[8] and prevents the induction of LTP deficits by Aβ oligomers in slices of entorhinal cortex[9], another memory centre. Other potential receptors for Aβ oligomers have been proposed, but we are unaware of studies that rival the evidence Laurén and colleagues present[4].

Laurén *et al.* find that, within PrP^C, amino-acid residues 95–110 are crucial for Aβ binding. Interestingly, the enzyme α-secretase — which precludes Aβ production by cleaving the Aβ precursor protein APP within the Aβ domain — also cleaves PrP^C between residues 111 and 112 (ref. 10), thus releasing from the membrane the portion of PrP^C to which Aβ would otherwise bind. So one way to prevent both Aβ production and the activation of downstream mediators by PrP^C might be to increase α-secretase activity.

Laurén and colleagues' observations create fertile ground for future investigations. One question is how the synthetic Aβ oligomers they characterized relate to the synthetic Aβ oligomers or natural, disease-associated oligomers that have been characterized by other researchers. For instance, does PrP^C mediate the effects of Aβ dimers isolated from brains of people with Alzheimer's disease[11], or of the Aβ*56 oligomer, which causes memory deficits in mouse models of this disease[1,2]?

How do the Aβ-induced effects on LTP that the authors observed in brain slices relate to cognitive impairments seen in Alzheimer's disease and in the related mouse models? To assess the clinical relevance and therapeutic potential of these findings[4], interaction between PrP^C and Aβ oligomers must be confirmed in patients with Alzheimer's disease, and the relationship between PrP^C levels and cognitive decline should be explored. It would also be of interest to determine whether the cognitive deficits and behavioural alterations seen in mouse models of Alzheimer's disease can be prevented by ablating or reducing PrP^C. For example, reducing the levels of the microtubule-associated

protein Tau prevents cognitive deficits in mice with high levels of Aβ oligomers[12]. Human Tau forms complexes with PrP (ref. 13), so the toxicity of Aβ oligomers might involve interactions between PrP^C and Tau.

And how exactly does the binding of Aβ to PrP^C affect neuronal plasticity? Do Aβ oligomers block, enhance or distort PrP^C functions? Do they exert their effects only on neurons or also on other brain cells such as microglia and astrocytes? Do they affect PrP^C directly or do they block interactions between PrP^C and a co-receptor (Fig. 1)? As Laurén *et al.* point out, the NMDA-type glutamate receptor interacts with PrP^C, modulating its function. PrP^C also activates a signalling cascade that regulates both NMDA receptors and Aβ-induced

synaptic deficits[14,15]. Alternatively, PrP^C may act as a 'Trojan horse', facilitating internalization of Aβ oligomers into vulnerable intracellular compartments (Fig. 1). Finally, PrP^C might accelerate Aβ oligomerization outside the cell or change the Aβ folding landscape inside by acting as a 'pathological chaperone'. These possibilities are not mutually exclusive.

Notwithstanding these unresolved questions, the discovery that PrP^C may be a mediator in the development of Alzheimer's disease is fascinating, not least from a therapeutic perspective. Drug development around this target should be facilitated by the already extensive knowledge of the cellular physiology of PrP^C and by the fact that PrP^C-deficient mice are viable and seem to have normal synaptic plasticity[5]. ■

Moustapha Cisse and Lennart Mucke are at the Gladstone Institute of Neurological Disease and in the Department of Neurology, University of California, San Francisco, San Francisco, California 94158, USA.
e-mail: lmucke@gladstone.ucsf.edu

1. Lesné, S. *et al. Nature* **440,** 352–357 (2006).
2. Cheng, I. H. *et al. J. Biol. Chem.* **282,** 23818–23828 (2007).
3. Shankar, G. M. *et al. J. Neurosci.* **27,** 2866–2875 (2007).
4. Laurén, J., Gimbel, D. A., Nygaard, H. B., Gilbert, J. W. & Strittmatter, S. M. *Nature* **457,** 1128–1132 (2009).
5. Aguzzi, A., Baumann, F. & Bremer, J. *Annu. Rev. Neurosci.* **31,** 439–477 (2008).
6. Prusiner, S. B. *Annu. Rev. Med.* **38,** 381–398 (1987).
7. Verdier, Y., Zarándi, M. & Penke, B. *J. Pept. Sci.* **10,** 229–248 (2004).
8. Sturchler, E., Galichet, A., Weibel, M., Leclerc, E. & Heizmann, C. W. *J. Neurosci.* **28,** 5149–5158 (2008).
9. Origlia, N. *et al. J. Neurosci.* **28,** 3521–3530 (2008).
10. Vincent, B., Cisse, M. A., Sunyach, C., Guillot-Sestier, M.-V. & Checler, F. *Curr. Alzheimer Res.* **5,** 202–211 (2008).
11. Shankar, G. M. *et al. Nature Med.* **14,** 837–842 (2008).
12. Roberson, E. D. *et al. Science* **316,** 750–754 (2007).
13. Wang, X.-F. *et al. Mol. Cell. Biochem.* **310,** 49–55 (2008).
14. Mouillet-Richard, S. *et al. Science* **289,** 1925–1928 (2000).
15. Chin, J. *et al. J. Neurosci.* **25,** 9694–9703 (2005).

ASTEROIDS

When planets migrate

Kevin J. Walsh

A study of the main asteroid belt, which lies between Mars and Jupiter, reveals an excess depletion of asteroids that was triggered by the migration of giant planets in the early history of the Solar System.

When astronomers study the asteroids confined to the main asteroid belt, they find that not only do these objects come in a variety of shapes and sizes, but also that their orbits can tell a tale about the evolution of the Solar System since the time planets formed. Although asteroids pale in size in comparison to planets, they share similar humble beginnings: in clouds of gas and dust circling the young Sun, grains of dust started sticking together, and soon there were boulders and then mountains of rock headed towards planethood. The newly formed planets were gravitationally dominant over asteroids, but

their orbits were not necessarily stationary. On page 1109 of this issue, Minton and Malhotra[1] take a close look at the main asteroid belt and show how the migration about 4 billion years ago of giant planets to the positions we see them in today affected asteroids' orbits.

Despite their large size, there is evidence that the giant planets — Jupiter, Saturn, Uranus and Neptune — have not been stationary since their formation[2,3]. Moreover, it is known that interactions with a disk of planetesimals (asteroids or comets consisting of rocks and ice left over from planet formation) can force planets to migrate inwards or outwards, shrinking or

50 & 100 YEARS AGO

50 YEARS AGO

A research station to study the effects of the aurora on radio reception is to be set up on the site of a former radar station at Hillhead, near Fraserburgh, in Aberdeenshire. This investigation, part of a programme of ionospheric studies, will be undertaken by the Stanford Research Institute of America in collaboration with scientists from the Royal Radar Establishment of the Ministry of Supply. Observations will be made with a large parabolic reflector about 140 ft. in diameter brought from America ... The Aberdeenshire site has been chosen because of its geographical suitability for the study of auroral phenomena. There will be no radiation danger, and interference with radio and television will be prevented by the operating conditions laid down for the installation.
From Nature 28 February 1959.

100 YEARS AGO

In the February number of *Naturen* Dr. L. Stejneger adduces further evidence in favour of the theory of the existence, at a comparatively recent date, of a land-bridge between Scotland and Scandinavia. This evidence is mainly based on the distribution of the species, or races, of charr (Salvelinus) ... *Salvelinus alpinus* is considered to be common to western Scandinavia and Scotland, while in eastern Scandinavia we have the typical *S. salvelinus* of the Alps. Iceland is the home of *S. nivalis*, while further north occur *S. insularis* and *S. stagnalis*. Lapland is the home of an intermediate form known as *S. salvelino-stagnalis*, while another annectant type, *S. alpino-stagnalis*, occurs in Greenland.
ALSO:
At the meeting of the Royal Society on Thursday, February 18, telegrams of congratulations on the hundredth anniversary of the birth of Charles Darwin were read from the University of Christiania, the University, Kharkoff ... the Society of Naturalists, Kharkoff ... and the Swedish Academy of Sciences, Stockholm.
From Nature 25 February 1909.
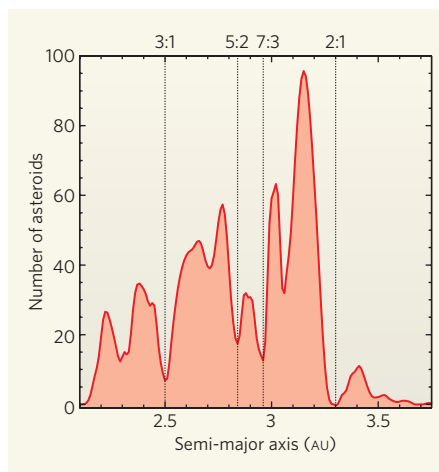
Figure 1 | Kirkwood gaps. The noticeable gaps in the distribution of asteroids as a function of the semi-major axes of their orbits (essentially their distances from the Sun) correspond to asteroid orbits whose periods are integer ratios of Jupiter's orbital period: 3 orbits for each 1 of Jupiter (located about 2.5 astronomical units (AU) from the Sun); 5 orbits for every 2 of Jupiter (2.82 AU); 7 orbits for every 3 of Jupiter (2.95 AU); 2 orbits for each 1 of Jupiter (3.27 AU).

expanding their orbits owing to the exchange of energy and angular momentum during close encounters with the planetesimals[4].

But planets have a much stronger effect on the planetesimals, and can destabilize their orbits throughout the asteroid belt. Looking at asteroid orbits in the main asteroid belt, as Daniel Kirkwood did in 1867 (ref. 5), we find big gaps, known as Kirkwood gaps. These gaps correspond to the location of orbital resonances with Jupiter — that is, of orbits whose periods are integer ratios of Jupiter's orbital period. For example, if an asteroid orbited the Sun 3 times for every time Jupiter did, it would be in a 3:1 orbital resonance with the planet, and would thus have an unstable orbit and eventually be cleared away from the asteroid belt (Fig. 1).

A consequence of planet migration is that the locations of planets' resonances also move, in turn affecting different parts of the asteroid belt at different times. Thus, if nothing has completely reshaped the asteroid belt since the planets settled into their current orbits, signatures of past planetary orbital migration may still remain. These are exactly what Minton and Malhotra[1] searched for.

In their study, the authors examined the same orbital parameter that Kirkwood did — the semi-major axis, which is essentially the size of the orbit, or distance from the Sun — for each of the 690 largest asteroids in the asteroid belt.

At the very least, this sample of asteroids has already been subject to the basic wear-and-tear — the occasional collision or ejection due to perturbations from the giant planets — of about 4 billion years of evolution after any planetary migration stopped. So Minton and Malhotra modelled the asteroid belt and followed the 4-billion-year evolution of the asteroids using a computer simulation that included the perturbing effects of the giant planets in their current orbits.

Comparison of the simulated asteroid population with the observed one revealed the expected Kirkwood gaps, but also a surprising feature: an excess depletion of asteroids in the observed population, both on the inner edge of the asteroid belt and on the outer edges of each Kirkwood gap. Qualitatively, it looks as if a snow plough were driven through the main asteroid belt, kicking out asteroids along the way and slowing to a stop at the inner edge of the belt.

Minton and Malhotra thus concluded that the observations could not be accounted for by considering the orbital evolution of the asteroid population solely under the gravitational tug of giant planets in their current orbits. But when the effects of the orbital migration of Jupiter and Saturn in the Solar System's early history were also taken into account, the simulations reproduced the observations quite well. In particular, the authors found that, with a relatively rapid migration of Jupiter inwards and the other giant planets outwards, the asteroid belt's population became substantially depleted as the resonances associated with each planet swept through it. The simulation's surviving population had a semi-major axis distribution similar to that observed, including the depletion

Figure 2 | The Moon's basins. The age of large basins on the Moon suggests that the Solar System went through a rapid period of 'late heavy bombardment' of impactors some 600 million years after planet formation. Minton and Malhotra's work[1] shows that the migration of giant planets about 4 billion years ago could have created plenty of potential impactors by removing asteroids from the main asteroid belt.

on the outer edges of the Kirkwood gaps. Thus, the migration of the giant planets seems to be able to account for the artefacts that Minton and Malhotra found in the observed distribution of the semi-major axes.

But, in addition to explaining the observed distribution of the semi-major axes, theories of planet migration will have to explain some of the more subtle features of the asteroid belt, specifically the distribution of eccentricities and inclinations of the asteroid orbits. The way in which these properties are altered by the sweeping resonances depends on the eccentricities and inclinations of the asteroids before migration takes place, so knowledge of the events preceding planet migration will be essential in explaining today's asteroid belt.

Planet migration may also be at the root of

the 'late heavy bombardment', an event believed to have occurred about 3.9 billion years ago — 600 million years after the birth of the Solar System — during which a large number of impactors hit the inner Solar System over a brief period of time. Evidence for this comes from the dating of basins on the Moon[6] (Fig. 2), which are plausibly the result of impacts from asteroids originating in the main asteroid belt[7]. And, indeed, Minton and Malhotra's work[1] shows that migrating giant planets can create plenty of potential impactors by removing asteroids from the asteroid belt. The issue is then one of matching up the timing of planet migration with the late heavy bombardment.

Minton and Malhotra's study provides an important step in our quest to understand the intricate evolution of the Solar System. This

process left the asteroid belt — and the larger Kuiper belt that lies beyond the planets — with an excessive depletion of asteroids, giant planets with eccentric and inclined orbits, and the Moon covered with impact basins. ■

Kevin J. Walsh is at the Observatoire de la Côte d'Azur, Laboratoire Cassiopée, BP 4229, 06304 Nice, France.
e-mail: kwalsh@oca.eu

1. Minton, D. A. & Malhotra, R. Nature 457, 1109–1111 (2009).
2. Malhotra, R. Nature 365, 819–821 (1993).
3. Malhotra, R. Astron. J. 110, 420–429 (1995).
4. Fernández, J. A. & Ip, W.-H. Icarus 58, 109–120 (1984).
5. Kirkwood, D. Meteoric Astronomy: A Treatise on Shooting-Stars, Fireballs, and Aerolites (Lippincott, 1867).
6. Tera, F., Papanastassiou, D. A. & Wasserburg, G. J. Earth Planet. Sci. Lett. 22, 1–21 (1974).
7. Strom, R. G., Malhotra, R., Ito, T., Yoshida, F. & Kring, D. A. Science 309, 1847–1850 (2005).

CLIMATE CHANGE

Southern see-saw seen

Jeffrey P. Severinghaus

The bipolar see-saw hypothesis provides an explanation for why temperature shifts in the two hemispheres were out of phase at certain times. The hypothesis has now passed a test of one of its predictions.

Abrupt climate changes during the last glacial period, about 110,000 to 10,000 years ago, have been widely documented in the Northern Hemisphere. But their underlying cause remains enigmatic. The most widely suspected culprit is the 'conveyor' circulation in the Atlantic Ocean, known to be capable of abruptly switching on and causing rapid warming of the European climate[1]. A clue comes from Antarctic temperature histories that show a roughly opposite pattern: when Europe warms abruptly, the south begins a gradual cooling[2]. This trading of hot and cold between the hemispheres has been called the bipolar see-saw[3], and it further implicates the Atlantic conveyor circulation because that circulation is known to orchestrate the redistribution of vast amounts of heat between the hemispheres.

A flaw in the case has long been the gradual, rather than abrupt, nature of the southern cooling. This has been variously explained as being due to the large heat reservoir represented by the Southern Ocean[4], which takes up to 1,000 years to cool off, or other inherently slow processes[5]. But the bipolar see-saw hypothesis predicts that, at the very least, the surface of the South Atlantic should respond abruptly in an antiphase fashion to the changes in the north.

On page 1097 of this issue, Barker and colleagues[6] present the first evidence of this rapid, opposite response, obtained from a sediment-core record of past temperature in the South Atlantic. The mark of a successful scientific theory is that it makes specific predictions that can be tested and found to be true. The passing

of this particular test, for the last deglaciation starting about 21,000 years ago, therefore strengthens the bipolar see-saw hypothesis.

Barker and colleagues also shed welcome light on the relationship between the concentration of carbon dioxide in the atmosphere and

the stability of ice sheets, a subject of intense current interest. A puzzle in palaeoclimatology has been the lag between the changes in Earth's orbit around the Sun that are thought to be the ultimate cause of the ends of ice ages (the Milankovitch hypothesis) and the catastrophic demise of the great ice sheets on North America and Scandinavia some 5,000–10,000 years later. Abundant evidence from glacial deposits and sea-level records[7] shows that the northern ice sheets began retreating 21,000–19,000 years ago, in agreement with Milankovitch theory, owing to the increasing amount of June sunshine falling on their margins at that time. Yet Antarctic temperature and CO_2 started rising only 18,000 years ago, and they did so rather abruptly, in contrast to the gradual, sinusoidal
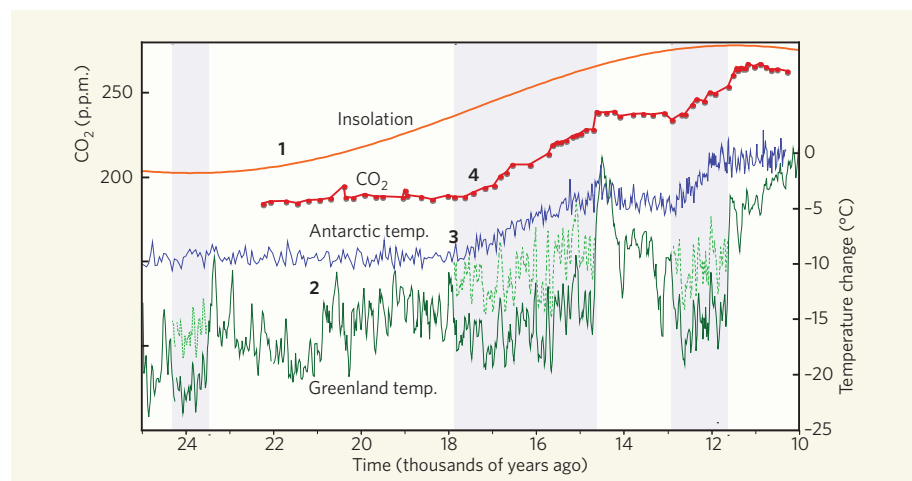

Figure 1 | Possible chain of events during the last deglaciation. 1, Starting about 24,000 years ago, a gradual change in Earth's spin-axis produced an increase in June insolation (sunshine) at 60° N. The northern ice sheets began to retreat, decreasing the planet's reflectivity and leading to warming (as recorded in Greenland; **2**) around 21,000–19,000 years ago. The resulting meltwater caused the North Atlantic conveyor to shut off around 18,000 years ago, switching the bipolar see-saw into its 'warm south' mode (shading) and producing abrupt warming of Antarctica (**3**) and the surrounding ocean. The consequent release of CO_2 from the deep ocean led to a rise in atmospheric CO_2 that began several hundred years later (**4**) — so warming the whole planet, keeping the meltwater coming, maintaining the 'warm south' mode, and increasing CO_2 even further in a positive feedback loop[8,9]. The upshot was destruction of the ice sheets. The Greenland temperature record is heavily influenced by winter temperature changes[14], so summer temperature is shown by adding 5 °C to the record during the 'see-saw south' intervals. (Data from refs 6, 8, 16 and 17.)

variation of the presumed orbital cause (Fig. 1). And it was only 14,000–10,000 years ago that the major (and apparently irreversible) collapse of the ice sheets occurred.

The interplay of the bipolar see-saw and CO_2 can provide a coherent explanation for these puzzles[8,9]. In this view, the abrupt onset of Antarctic warming 18,000 years ago can be attributed to the bipolar see-saw, due to a switch-off of the Atlantic circulation in response to the crossing of some threshold by the increasing injection of meltwater from the northern ice sheets. Hard on the heels of this warming, about 300 years later[10], CO_2 levels in the atmosphere started to rise as the deep ocean warmed and released CO_2 owing to decreased stratification[11] or decreased Antarctic sea ice[12]. This increase is in accord with observations of a simultaneous rise in the abundances of the associated krypton and xenon markers in the atmosphere[13].

Carbon dioxide then warmed the entire planet through the greenhouse effect and contributed to further melting of the great ice sheets. This melting ice in turn kept the bipolar see-saw in its 'warm south' polarity in a positive feedback loop, wherein the sustained supply of meltwater ensured continued release of CO_2 and thereby continued warming and melting[8,9]. Extreme seasonality owing to winter sea-ice cover on the North Atlantic explains why Greenland's mean annual temperature records seem cold at this time, despite warm summers and abundant melting[14].

Only when the bipolar see-saw switched back to its 'warm north' polarity 14,700 years ago, perhaps due to de-densification of the deep ocean caused by the Antarctic's warming and the addition of fresh water from rain and melting ice[15], did the increase of CO_2 come to a temporary halt. At this point, the triple whammy of switched-on Atlantic Ocean circulation in the presence of high CO_2 and strong June sunshine provided the *coup de grâce* for the northern ice sheets.

If this hypothesized link between the bipolar see-saw and CO_2 is supported by further research, there are implications of potential relevance to future climate. First, it is difficult to explain the demise of the ice sheets without the added heating from CO_2, confirming that this gas has killed ice sheets in the past and may do so again. Second, the predicted slowdown of Atlantic circulation in the coming century may cause an additional release of CO_2 from the ocean that adds to the human-made CO_2, a biogeochemical feedback that is not considered in current climate projections.

To sum up, the chain of causality for the last deglaciation began with Earth's spin-axis. But this gradual change was amplified in an unsteady cascade of events by ice cover, meltwater, ocean circulation and CO_2 in the atmosphere. Barker and colleagues' identification[6] of the see-saw in records of past climate helps make sense of the sometimes asynchronous sequence of changes in these disparate phenomena. ■

Jeffrey P. Severinghaus is at the Scripps Institution of Oceanography, La Jolla, California 92093-0244, USA.
e-mail: jseveringhaus@ucsd.edu

1. Broecker, W. S. & Denton, G. H. *Geochim. Cosmochim. Acta* **53,** 2465–2501 (1989).
2. Blunier, T. & Brook, E. J. *Science* **291,** 109–112 (2001).
3. Broecker, W. S. *Paleoceanography* **13,** 119–121 (1998).
4. Stocker, T. F. & Johnsen, S. J. *Paleoceanography* doi:0.1029/2003PA000920 (2003).
5. Keeling, R. F. & Visbeck, M. *Quat. Sci. Rev.* **24,** 1809–1820 (2005).
6. Barker, S. *et al. Nature* **457,** 1097–1102 (2009).
7. Yokoyama, Y., Lambeck, K., DeDeckker, P., Johnston, P. & Fifield, L. K. *Nature* **406,** 713–716 (2000).
8. Paillard, D. *Rev. Geophys.* **39,** 325–346 (2001).
9. Denton, G. H. *et al. Pages News* **14,** 14–16 (2006).
10. Ahn, J. & Brook, E. J. *Science* **322,** 83–85 (2008).
11. Toggweiler, J. R. *Paleoceanography* **14,** 571–588 (1999).
12. Stephens, B. B. & Keeling, R. F. *Nature* **404,** 171–174 (2000).
13. Headly, M. A., Kawamura, K. & Severinghaus, J. P. *Eos Trans. AGU* **89**(53), Fall Mtg Suppl., Abstr. PP41F-04 (2008).
14. Denton, G. H., Alley, R. B., Comer, G. C. & Broecker, W. S. *Quat. Sci. Rev.* **24,** 1159–1182 (2005).
15. Knorr, G. & Lohmann, G. *Nature* **424,** 532–536 (2003).
16. Jouzel, J. *et al. Science* **317,** 793–797 (2007).
17. Cuffey, K. M. & Clow, G. D. *J. Geophys. Res.* **102,** 26383–26396 (1997).

PALAEONTOLOGY

Birth of the jawed vertebrates

Per E. Ahlberg

The discovery of embryos in certain fossil fishes not only shows that internal fertilization and live birth evolved early in vertebrate history, but also raises questions about the origin of jawed vertebrates.

Every once in a while, a discovery comes along that puts our biological understanding of some extinct group of organisms on a much firmer footing. On page 1124 of this issue, Long and colleagues[1] present such a discovery, one that may prove to have far-reaching implications for our understanding of early vertebrate evolution. For the second time in less than a year[2], they have found preserved embryos in the body cavity of a placoderm fish.

Placoderms are extinct jawed fishes that lived from approximately 430 million years ago to the end of the Devonian 360 million years ago. Their heads and shoulder girdles are covered in bony armour; plates of bone also form the biting surfaces of the jaws, sometimes flat and covered
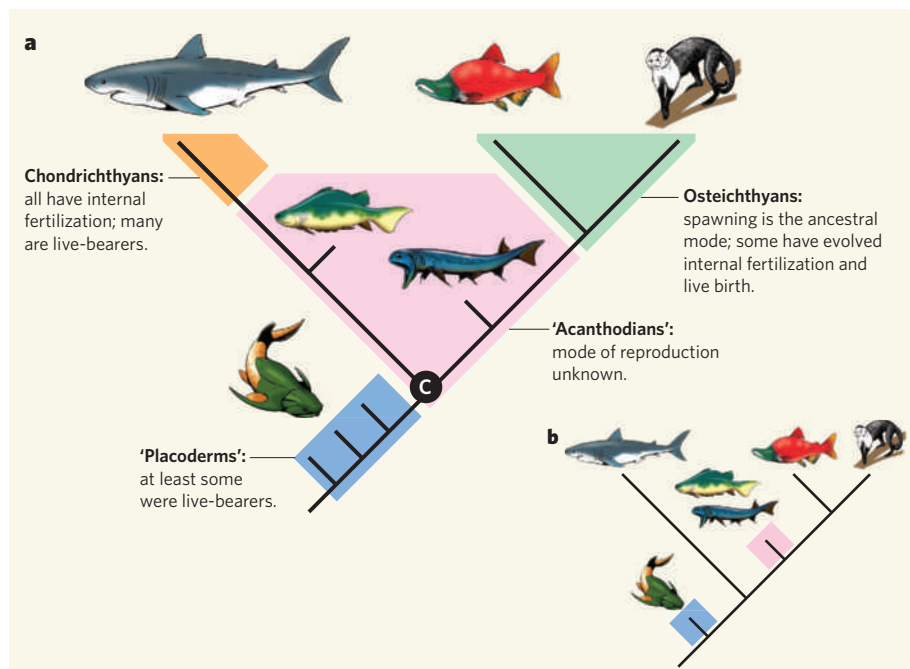
Figure 1 | Gnathostomes and their modes of reproduction. a, Reproductive modes mapped onto a gnathostome phylogeny, simplified from ref. 7. 'C' marks the gnathostome crown-group node: the part of the tree that lies below this node is the gnathostome stem group; that above it is the crown group. In this scheme, both osteichthyans and chondrichthyans are clades (groups comprising all descendants of a single common ancestor), but the 'placoderms' and 'acanthodians' are not — hence the inverted commas. Examples of placoderm groups are the antiarchs, ptyctodonts and arthrodires, the latter including *Incisoscutum*[1] (the antiarchs are placed as the lowest placoderm branch and the arthrodires as the highest in ref. 7). **b,** A conventional consensus phylogeny: placoderms and acanthodians are interpreted as clades. (Animal images by M. D. Brazeau, reproduced with permission.)

OBESITY
Fat chance

Eat more, move less and lose weight — sounds too good to be true. Yet this is exactly what Julia Fischer and her colleagues observed in mice lacking one specific gene (J. Fischer et al. Nature advance online publication doi:10.1038/nature07848; 2009).

Obesity is a complex disorder because, as well as environmental factors, many genes seem to be involved. One such gene is FTO, as several studies have indicated that different versions of FTO are strongly correlated with body mass index: individuals carrying the high-risk version weigh roughly 3 kilograms more than those with the low-risk version.

Fischer et al. studied mice lacking Fto (Fto[−/−] mice) and compared them with normal mice and with those carrying only one copy of the gene (Fto[+/−] mice). The absence of Fto did not affect embryonic development, but by six weeks after birth, Fto[−/−] mice weighed on average 30–40% less than normal or Fto[+/−] mice. This reduction in weight was associated with a marked loss of white fat tissue, with near-complete loss by 15 months.

The lower weight of the mutant mice doesn't seem to be due to reduced calorific intake. In fact, these mice ate more in proportion to their body weight than normal mice. Moreover, on a high-fat diet, both groups of mutant mice gained much less weight than normal animals. Instead, Fto[−/−] mice used more energy, while not moving much. The authors suggest that this increased energy expenditure might be due to higher activity of the sympathetic nervous system — that is, to enhanced circulating levels of adrenaline and noradrenaline.

Fischer and colleagues' data indicate that variations in the human FTO gene might affect its levels of expression, either putting individuals at risk of obesity or protecting them from it. It remains to be seen how FTO might regulate the activity of the sympathetic nervous system.

Sadaf Shadan

with small bumps, sometimes developed into nasty, self-sharpening scissor blades. The largest of these fishes equalled the size of a great white shark and must have been formidable predators, but most were less than a metre in length.

Working with three-dimensional placoderm fossils from Gogo in Western Australia, Long and colleagues have discovered specimens of three different placoderms, Materpiscis[2], Austroptyctodus[2] and Incisoscutum[1], that contain minute but perfectly preserved armour plates of the same species in their body cavities. When first discovered, they were thought to be stomach contents[3]. But the plates show no bite marks or etching by stomach acids, and are not mixed with bones from other species; they are the remains of unborn embryos. In Materpiscis, a curving tubular structure associated with one of them has been interpreted as an umbilical cord[2].

Embryos in the body cavity imply internal fertilization. It was noted long ago[4] that ptyctodonts, the placoderm subgroup to which Materpiscis and Austroptyctodus belong, have sexually dimorphic pelvic fins, somewhat like the 'claspers' used for internal fertilization in sharks. Arthrodires, the placoderm group that includes Incisoscutum, lack the sexually dimorphic external bones present on the pelvic fins of ptyctodonts. However, Long et al. argue that the partially preserved internal fin skeletons of their specimens indicate a shark-like structure, probably implying sexual dimorphism and internal fertilization. Ptyctodonts and arthrodires seem to be closely related, and so internal fertilization, and possibly live birth of young, are probably shared features retained from their common ancestor.

The living jawed vertebrates, or gnathostomes, fall into two groups, the Chondrichthyes and the Osteichthyes (Fig. 1). The Chondrichthyes (sharks, rays and ratfishes) all have internal fertilization, and many give birth to live young, whereas the ancestral condition for the Osteichthyes (ray-finned fishes, lobe-finned fishes and land vertebrates) is to spawn small eggs that are fertilized externally. Live-bearers tend to produce much fewer young than external spawners and have lower potential rates of population growth. This contrast in reproduction puts a new perspective on the ecology of the Gogo environment, a tropical reef[5], where a wide diversity of placoderms coexisted with lungfishes and primitive ray-finned fishes that were probably externally fertilizing spawners. It is also interesting to note that the extinction of the placoderms at the end of the Devonian was followed by a major diversification of chondrichthyans. But it is to the study of gnathostome interrelationships that the discoveries of Long et al. may prove to be most pertinent.

Ideas about the origin of gnathostomes are currently in a state of flux. For much of the twentieth century, placoderms were regarded as relatives or possibly ancestors of chondrichthyans[4], partly because they seemed to use internal fertilization. But recently the majority view has placed them in the gnathostome stem group[6] — that is, the common ancestral lineage of the living jawed vertebrates. A new analysis by Brazeau[7] suggests that placoderms may not be a natural group at all, but a 'paraphyletic array' spread out along the gnathostome stem (Fig. 1a; contrast with Fig. 1b). If that is correct, placoderms become extremely informative about the origin of jawed vertebrate morphology. This is where the evidence for internal fertilization and live-bearing in placoderms becomes important.

The ancestral mode of reproduction for osteichthyans seems to be external fertilization. The distribution of live-bearing among living vertebrates strongly suggests that internally fertilizing live-bearers are unlikely to give rise to externally fertilizing spawners, so we would not expect the osteichthyan stem lineage, or the gnathostome stem lineage below it, to contain a segment characterized by live-bearing. Brazeau's analysis[7] places the ptyctodonts and arthrodires as successive branches off the gnathostome stem, implying the existence of such a segment unless the two groups have evolved live-bearing independently. However, only a minor change in the tree would be needed to join ptyctodonts and arthrodires together in a clade (that is, forming a single side branch), and thus make the offending stem segment disappear. A more important question is whether the most primitive placoderms, such as the antiarchs (bottom-feeding fishes with armoured pectoral fins), were also live-bearers, because this would undermine the case for the placoderms forming a paraphyletic segment of the gnathostome stem.

Long and colleagues[1] argue that the antiarchs had external fertilization. They lack pelvic fins altogether, and fossils have been found of free-living juveniles that are small and undeveloped enough to correspond to the embryos of Materpiscis, Austroptyctodus and Incisoscutum. It may thus be that both internal fertilization and live-bearing evolved within the placoderms. Perhaps this was a unique innovation in one placoderm clade. Alternatively, could some placoderms be stem gnathostomes and others, those with internal fertilization, stem chondrichthyans? Possibly, but this conflicts with new evidence that the acanthodians (vaguely shark-like fishes, with fin spines and tiny scales, which became extinct about 250 million years ago) form a paraphyletic array encompassing the bases of the chondrichthyan and osteichthyan lineages (Fig. 1a)[6].

The tangled skein of jawed-vertebrate origins continues to challenge researchers. But discoveries such as the placoderm embryos of Gogo are giving us the tools to gradually untangle it — as well as showing us intimate glimpses of life in a lost world. ■

Per E. Ahlberg is in the Subdepartment of Evolutionary Organismal Biology, Department of Physiology and Developmental Biology, Uppsala University, 752 36 Uppsala, Sweden.
e-mail: per.ahlberg@ebc.uu.se

1. Long, J. A., Trinajstic, K. & Johanson, Z. Nature 457, 1124–1127 (2009).
2. Long, J. A., Trinajstic, K., Young, G. C. & Senden, T. Nature 453, 650–653 (2008).
3. Dennis, K. & Miles, R. S. Zool. J. Linn. Soc. 73, 213–258 (1981).
4. Moy-Thomas, J. A. & Miles, R. S. Palaeozoic Fishes (Chapman & Hall, 1971).
5. Long, J. A. Swimming in Stone: The Amazing Gogo Fossils of the Kimberley (Fremantle, 2008).
6. Janvier, P. Early Vertebrates (Oxford Sci. Publ., 1996).
7. Brazeau, M. D. Nature 457, 305–308 (2009).

Is REST required for ESC pluripotency?

Arising from: S. K. Singh, M. N. Kagalwala, J. Parker-Thornburg, H. Adams & S. Majumder *Nature* **453, 223–227 (2008)**

The DNA-binding protein REST (also called NRSF) is a transcriptional repressor that targets many neuronal genes[1,2] and is abundant in human and mouse pluripotent embryonic stem cells (ESCs)[3–5]. In a recent Letter to *Nature*, Singh et al.[6] suggested that REST controls the self-renewal and pluripotency of ESCs, because they found that ESCs in which a single REST allele was disrupted (Fig. 1a, β-geo-stop insertion) had reduced alkaline phosphatase activity and expressed lower levels of several pluripotency-associated genes[6]. Here we show that partial or complete loss of functional REST protein does not abrogate ESC potential as reflected by marker gene expression. These data are consistent with earlier reports[7,8], and argue that REST is not required for maintaining ESC pluripotency.

Heterozygous *Rest*-targeted ESCs (Fig. 1a) that were isolated from a previously published mouse line[7] displayed a reduced level of full-length REST, and no full-length REST protein was observed in homozygous *Rest* mutant ESCs (Fig. 1b). The truncated REST peptide encoded by the mutated allele lacks DNA-binding capacity[7] and fails to bind REST target genes, as demonstrated by chromatin immuno-precipitation (Fig. 1c). We detected no substantial change in Oct4 (also called Pou5f1) protein levels or alkaline phosphatase activity in heterozygous ($+/-$) or homozygous ($-/-$) *Rest* ESCs relative to matched wild-type ($+/+$) controls (Fig. 1b, d). Analysis by quantitative polymerase chain reaction after reverse transcription (RT–PCR) demonstrated that the transcript levels of the pluripotency markers *Oct4*, *Nanog* and *Sox2* were similar in wild-type, *Rest*$^{+/-}$ and *Rest*$^{-/-}$ ESCs (Table 1). The REST target genes *Syp* and *Syt4* were de-repressed in the absence of REST. In contrast, genes that are induced on differentiation to endoderm (*Gata4*), mesoderm (*Bry*, also known as *T*) or ectoderm (*Sox1*) were expressed at low levels in both wild-type and *Rest* mutant ESCs compared to differentiated control tissues (Table 1). Hence, the expression of pluripotency markers and differentiation-associated genes was unaffected in REST-deficient ESCs. This is in contrast to the results reported by Singh et al.[6], but it is consistent with the phenotype of two independent, publicly available *Rest*$^{+/-}$ ESC lines (Fig. 1e, f). To address the possibility that the discrepancy between their results[6] and ours reflects either differences in the targeting approaches or selection for secondary events that compensate for REST deficiency, we performed REST knockdown experiments in feeder-independent wild-type ESCs using culture conditions and protocols similar to those described by Singh et al.[6]. The RNA interference analysis confirmed that markers of pluripotency and differentiation are not reproducibly deregulated in ESCs depleted of REST by either short hairpin RNA (shRNA) or short interfering RNA (siRNA) knockdown (Fig. 1b and Table 1, marked by asterisks).

Our results do not rule out that REST mutant ESCs may have a subtle phenotype; however, the conclusion by Singh et al.[6] that REST is required to maintain ESC self-renewal and pluripotency appears inconsistent with the results presented here.
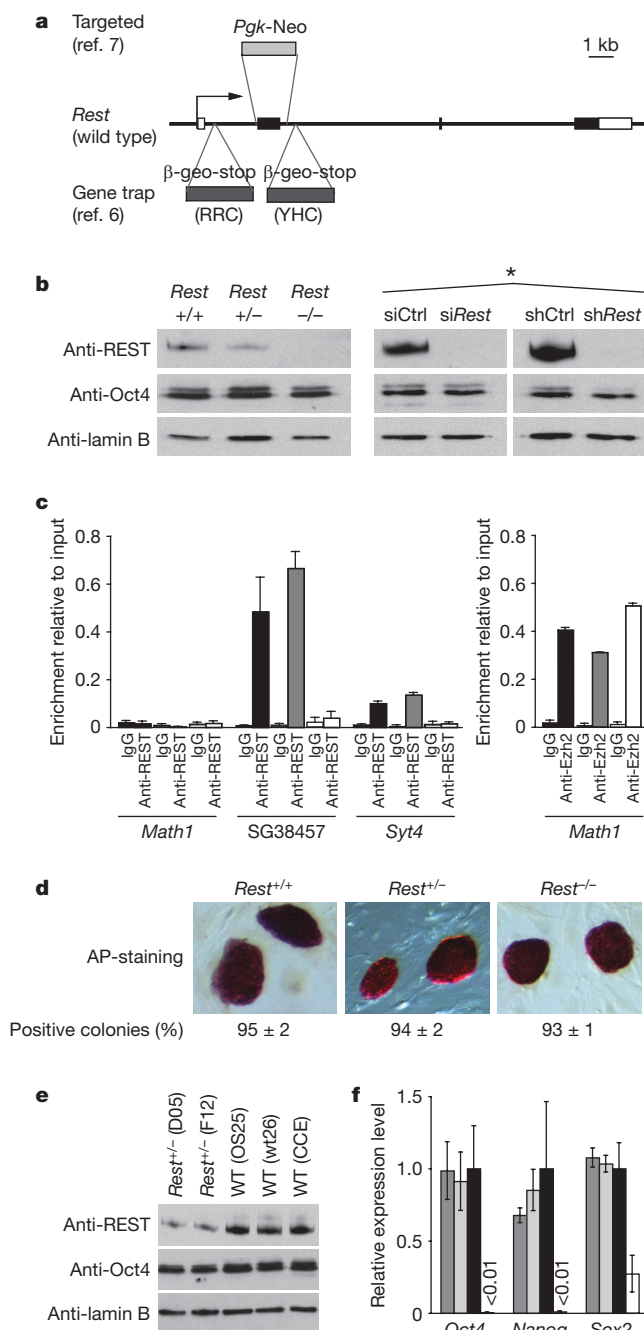
Figure 1 | ESCs deficient in REST retain stem cell characteristics. a, Genetic alterations to the *Rest* locus analysed here (phosphoglycerate kinase 1-neomycin resistance gene (*Pgk*-Neo) insertion[7]) and by Singh et al.[6] (β-geo-stop insertions in the gene trap ESC lines RRC and YHC). Rectangles represent exons; coding regions are in black. **b**, REST, Oct4 and lamin B protein levels in wild-type ($+/+$), homozygous ($-/-$) and heterozygous ($+/-$) *Rest* ESC lysates, and (marked by an asterisk) in wild-type ESCs transfected with siRNA or shRNA constructs targeting *Rest* (si*Rest* and sh*Rest*) or a control sequence (siCtrl and shCtrl). **c**, Chromatin immunoprecipitation of *Rest*$^{+/+}$ (black), *Rest*$^{+/-}$ (grey) and *Rest*$^{-/-}$ (white) ESCs using anti-REST (left panel) and anti-Ezh2 (right panel) versus control antibody (immunoglobulin G, IgG). REST binds SG38457 (also known as *Fam70b*) and *Syt4* (which both contain a RE1 motif) but not *Math1* (also known as *Atoh1*; RE1-negative). No significant binding of REST was detected in *Rest*$^{-/-}$ ESCs. Ezh2-binding at *Math1* (ref. 10) confirmed that chromatin fragments were intact. Error bars represent the standard deviation of three experiments. **d**, Alkaline phosphatase (AP) activity of mutant ESCs (percentage AP-positive colonies ± standard deviation). **e**, REST, Oct4 and lamin B protein levels in two additional *Rest*$^{+/-}$ (D05 and F12) and three wild-type ESC lysates. **f**, Transcript levels in D05 (dark grey), F12 (light grey) and wild-type (black) ESCs are shown (and, for comparison, retinoic-acid-treated embryoid bodies (white)) relative to wild type. Values were normalized to housekeeping genes and error bars show standard deviation of 4–6 experiments.

E4

Table 1 | Gene expression by REST-deficient ESCs

	Rest	Oct4	Nanog	Sox2	Transcript levels† Syp	Syt4	Gata4	Bry	Sox1
Rest[+/+]	100	100	100	100	100	100	100	100	100
Rest[+/−]	62 ± 0.4	77 ± 30	48 ± 11	87 ± 13	208 ± 91	64 ± 52	76 ± 15	118 ± 47	157 ± 44
Rest[−/−]	1 ± 0.5	96 ± 13	82 ± 18	95 ± 3	8,356 ± 2,678	1,291 ± 327	147 ± 19	337 ± 106	93 ± 13
Controls‡	9 ± 5	<1	<1	21 ± 10	10,000	2,700	1,500	2,300	2,000
siCtrl*	100	100	100	100	100	100	100	100	100
siRest*	15 ± 2	107 ± 5	143 ± 26	96 ± 6	1,840 ± 279	336 ± 58	96 ± 14	128 ± 28	114 ± 24
shCtrl*	100	100	100	N.A.	100	100	100	100	100
shRest*	17 ± 7	85 ± 0.1	86 ± 6	N.A.	1,602 ± 39	352 ± 109	91 ± 10	77 ± 26	91 ± 17

* RNAi-mediated REST knockdown experiments.
† Transcipt levels were detected by real-time RT–PCR (primers available on request) and values were normalized to housekeeping genes (*Hmbs, Gapdh, Ubc*). Results are shown relative to wild-type ESCs (*Rest*[+/+]) or control ESC samples (siCtrl, shCtrl). The average and standard deviation of 3–6 independent experiments are displayed.
‡ Differentiated ESCs (retinoic-acid-treated embryoid bodies) provide negative controls for *Rest, Oct4, Nanog* and *Sox2*, whereas detection of *Syp, Syt4, Sox1* (in E15 embryo head), *Gata4* (in E16 embryo liver) and *Bry* (in ES-cell-derived mesoderm) provide positive controls.

METHODS

Wild-type, *Rest*[+/−] and *Rest*[−/−] ESCs were derived from mutant *Rest* animals[7] by standard methods[9] and cultured on mitotically inactivated embryonic fibroblasts in the presence of leukaemia inhibitory factor (LIF, 1,000 U ml[−1]). No difference in the phenotype of early passage (P5–P8) and late passage (>P30) cells was observed (data not shown). Two independent *Rest*[+/−] ESC lines (D034D05 (D05) and D034F12 (F12)) were obtained from the German Genetrap Consortium and cultured as described above. Culturing of wt26 and CCE wild-type ESC lines has been described[10,11]. E14Tg2A-derived 46C (ref. 12) and OS25 (ref. 13) ESC lines were cultured in the presence of LIF on gelatinized plates throughout. 46C ESCs were transfected with siRNA targeting REST or with negative control siRNA and analysed two days later. OS25 ESCs were transfected with shRNA constructs targeting REST or a control sequence and collected after three days as described previously[11]. Alkaline phosphatase staining, western blotting, chromatin immunoprecipitation and gene expression analyses were performed as described[10,11].

Helle F. Jørgensen[1]**, Zhou-Feng Chen**[2]**, Matthias Merkenschlager**[1] **& Amanda G. Fisher**[1]

[1]Lymphocyte Development Group, MRC Clinical Sciences Centre, Imperial College School of Medicine, Hammersmith Hospital Campus, Du Cane Road, London, W12 0NN, UK.
e-mail: helle.jorgensen@imperial.ac.uk
e-mail: amanda.fisher@csc.mrc.ac.uk
[2]Departments of Anesthesiology, Psychiatry and Developmental Biology, Washington University School of Medicine Pain Center, Saint Louis, Missouri 63110, USA.

1. Chong, J. A. *et al.* REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80**, 949–957 (1995).
2. Schoenherr, C. J. & Anderson, D. J. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* **267**, 1360–1363 (1995).
3. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
4. Westbrook, T. F. *et al.* SCFβ-TRCP controls oncogenic transformation and neural differentiation through REST degradation. *Nature* **452**, 370–374 (2008).
5. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049–1061 (2008).
6. Singh, S. K., Kagalwala, M. N., Parker-Thornburg, J., Adams, H. & Majumder, S. REST maintains self-renewal and pluripotency of embryonic stem cells. *Nature* **453**, 223–227 (2008).
7. Chen, Z. F., Paquette, A. J. & Anderson, D. J. NRSF/REST is required *in vivo* for repression of multiple neuronal target genes during embryogenesis. *Nature Genet.* **20**, 136–142 (1998).
8. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38**, 431–440 (2006).
9. Chen, Z. F. & Behringer, R. R. twist is required in head mesenchyme for cranial neural tube morphogenesis. *Genes Dev.* **9**, 686–699 (1995).
10. Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nature Cell Biol.* **8**, 532–538 (2006).
11. Jorgensen, H. F. *et al.* The impact of chromatin modifiers on the timing of locus replication in mouse embryonic stem cells. *Genome Biol.* **8**, R169 (2007).
12. Conti, L. *et al.* Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.* **3**, e283 (2005).
13. Billon, N., Jolicoeur, C., Ying, Q. L., Smith, A. & Raff, M. Normal timing of oligodendrocyte development from genetically engineered, lineage-selectable mouse ES cells. *J. Cell Sci.* **115**, 3657–3665 (2002).

Is REST a regulator of pluripotency?

Arising from: S. K. Singh, M. N. Kagalwala, J. Parker-Thornburg, H. Adams & S. Majumder *Nature* **453**, 223–227 (2008)

Establishment and maintenance of the pluripotent state of ESCs is a key issue in stem cell biology and regenerative medicine, and consequently identification of transcription factors that regulate ESC pluripotency is an important goal. Singh *et al.*[1] claim that the transcriptional repressor REST is such a regulator and that a 50% reduction of REST in ESCs leads to activation of a specific microRNA, miR-21, and that this subsequently results in loss of pluripotency markers and a reciprocal gain in some lineage-specific differentiation markers. In contrast, we show that, in haplodeficient *Rest*[+/−] ESCs, we detected no change in pluripotency markers, no precocious expression of differentiated neuronal markers and no interaction of REST with miR-21. It is vital that identification of factors that regulate pluripotency is based on robust, consistent data, and the contrast in data reported here undermines the claim by Singh *et al.*[1] that REST is such a regulator.

Singh *et al.*[1] use mouse ESCs in which one allele of *Rest* was knocked out by insertion of gene traps (YHC334 and RRC160 cells) and by use of *Rest* siRNA. Both manipulations result in a 50% loss of REST protein. Using these two experimental systems, Singh *et al.*[1] then go on to characterize expression of pluripotent and lineage-specific markers in self-renewing E14 ESCs, and during their differentiation into embryoid bodies. They observe that decreased *Rest* levels lead to a loss of pluripotency markers such as Oct4 and Nanog, and an increase in expression of lineage-specific differentiation markers including induction of neuronal progenitor markers Ngn2 and Mash1. Most significantly, the authors conclude that these changes are a consequence of direct transcriptional repression of a microRNA, miR-21.

However, these conclusions are in marked contrast to previous results that have shown that RNA-interference-mediated reduction of *Rest* levels in ESCs does not induce loss of pluripotency markers[2].
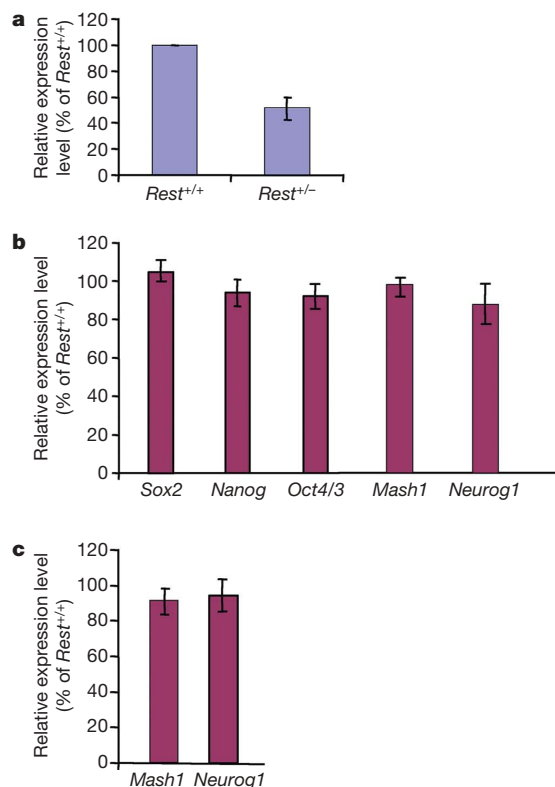
Figure 1 | *Rest*$^{+/-}$ ESCs retain pluripotent markers and show no precocious induction of neural marker genes. a, Relative expression levels of *Rest* in wild-type (*Rest*$^{+/+}$) and haplodeficient (*Rest*$^{+/-}$) mouse ESCs assessed using real-time PCR analysis. **b,** Relative levels of pluripotency genes *Sox2*, *Nanog* and *Oct4/3* and neuronal progenitor genes *Mash1* and *Neurog1* in undifferentiated *Rest*$^{+/-}$ compared to wild-type *Rest*$^{+/+}$ cells. No significant differences were detected using Student's two-tailed *t*-test at $P < 0.05$. **c,** Relative levels of neuronal progenitor genes *Mash1* and *Neurog1* in *Rest*$^{+/-}$ compared to wild-type *Rest*$^{+/+}$ cells after four days of monolayer differentiation. No significant differences were detected using Student's *t*-test at $P < 0.05$. Error bars indicate standard error of the mean of three independent experiments.

Furthermore, loss of either one or both *Rest* alleles does not lead to any discernible dysfunction in gastrulation in transgenic mice[3]. Here we present additional evidence, using ESCs with one allele of *Rest* deleted (*Rest*$^{+/-}$), that argues against a role of REST in regulating pluripotency genes. In our *Rest*$^{+/-}$ ESCs, despite a 50% reduction in *Rest* levels (Fig. 1a), we see no change in expression levels of *Sox2*, *Nanog* or *Oct4/3* between wild-type ESCs or in *Rest*$^{+/-}$ cells (Fig. 1b), nor do we observe precocious induction of neuronal progenitor markers *Mash1* and *Neurog1* (Fig. 1c). In support of these findings, DNA microarray analysis of ESCs treated with either a dominant-negative REST construct[4] or siRNA (data not shown) does not show downregulation of any pluripotency genes.

The second strand of data in Singh *et al.*[1] pertains to the repression of miR-21 by REST, and the suggestion that it mediates the pluripotency function of REST in ESCs through its repression of transcripts of *Nanog*, *Sox2*, *Tbx3* and *c-Myc*. Several facts argue against this interpretation. First, neither ChIP-PET (ChIP paired-end diTag)

nor ChIP-PCR analysis of E14 mouse ESC chromatin provides any evidence for recruitment of REST at the miR-21 locus[4]. This absence of REST binding is also evident in other genome-wide ChIP analyses[5,6]. Second, another study showed that in human ESCs miR-21 levels actually drop during embryoid body differentiation[7] arguing against a role for miR-21 in the downregulation of pluripotency markers and contradicting the data of Singh *et al.*[1], who report an increase in miR-21 levels in embryoid bodies compared with ESCs. Third, REST does not regulate levels of miR-21 in a transformed embryonic striatal cell line[8] where both are robustly expressed[9]. Although, there may be some cell and/or context specificity to REST occupancy, most genome-wide ChIP studies show a remarkable overlap of target genes and, as yet, no such study has indicated miR-21 as a REST target gene. Further, we emphasize that, as with Singh *et al.*[1], all our data are derived using feeder-free conditions[4].

In light of these conflicting lines of diverse evidence from multiple laboratories, we feel that, at present, it is premature to conclude that REST "is a newly discovered element of the interconnected regulatory network that maintains the self-renewal and pluripotency of mouse ES cells"[1].

METHODS

A conditional knockout targeting vector was used to flox one of the *Rest* alleles in HM1 ESCs followed by transient expression of *Cre* recombinase to delete the floxed allele. ESCs were differentiated into neural progenitor cells and subsequently to mature neurons using a monolayer culture in N2B27 medium[10]. Gene expression analysis was carried out using real-time PCR and data analysed using Student's *t*-test and $P < 0.05$ and $P < 0.01$. ChIP-PET analysis[11] of chromatin was carried out as described using anti-REST antibody (Upstate 07-579). All ESCs were cultured feeder-free; full details can be found elsewhere[4].

Noel J. Buckley[1]**, Rory Johnson**[2]**, Yuh-Man Sun**[1] **& Lawrence W. Stanton**[2]
[1]King's College London, Institute of Psychiatry, Centre for the Cellular Basis of Behaviour, James Black Centre, 125 Coldharbour Lane, London SE5 9NU, UK.
e-mail: noel.buckley@iop.kcl.ac.uk
[2]Stem Cell and Developmental Biology, Genome Institute of Singapore, 60 Biopolis Street #02-01, Genome Building, Singapore 138672.

1. Singh, S. K., Kagalwala, M. N., Parker-Thornburg, J., Adams, H. & Majumder, S. REST maintains self-renewal and pluripotency of embryonic stem cells. *Nature* **453**, 223–227 (2008).
2. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38**, 431–440 (2006).
3. Chen, Z. F., Paquette, A. J. & Anderson, D. J. NRSF/REST is required *in vivo* for repression of multiple neuronal target genes during embryogenesis. *Nature Genet.* **20**, 136–142 (1998).
4. Johnson, R. *et al.* REST regulates distinct transcriptional networks in embryonic and neural stem cells. *PLoS Biol.* **6**, e256 (2008).
5. Otto, S. J. *et al.* A new binding motif for the transcriptional repressor REST uncovers large gene networks devoted to neuronal functions. *J. Neurosci.* **27**, 6729–6739 (2007).
6. Johnson, D., Mortazavi, A., Myers, R. & Wold, B. Genome-wide mapping of *in vivo* protein–DNA interactions. *Science* **316**, 1497–1502 (2007).
7. Morin, R. *et al.* Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* **18**, 610–621 (2008).
8. Zuccato, C. *et al.* Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nature Genet.* **35**, 76–83 (2003).
9. Johnson, R. *et al.* A microRNA-based gene dysregulation pathway in Huntington's disease. *Neurobiol. Dis.* **29**, 438–445 (2008).
10. Ying, Q. L. & Smith, A. G. Defined conditions for neural commitment and differentiation. *Methods Enzymol.* **365**, 327–341 (2003).
11. Wei, C. L. *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207–219 (2006).

Singh et al. reply

Replying to: H. F. Jørgensen, Z.-F. Chen, M. Merkenschlager & A. G. Fisher *Nature* **457,** doi:10.1038/nature07783; N. J. Buckley, R. Johnson, Y.-M. Sun & L. W. Stanton *Nature* **457,** doi:10.1038/nature07784 (2009)

In contrast to the comments made by Jørgensen et al.[1] and Buckley et al.[2], our experiments showed that REST maintains the self-renewal and pluripotency of mouse embryonic stem cells (mESCs)[3]. Two recent papers support our work: ref. 4 indicated that REST is indeed in the network that regulates ESC self-renewal and pluripotency and ref. 5 showed that mESCs with lower REST levels derived from a mouse model of Down's syndrome have decreased levels of self-renewal markers and a higher propensity towards differentiation, even when cultured in the presence of LIF. We note that Buckley and Stanton also recently concluded that REST is part of the Oct4–Sox2–Nanog regulatory network and has "a key role in the maintenance of the ESC phenotype"[6]. We proposed that REST represses a set of microRNAs that potentially target self-renewal genes. At least one of them, miR-21, represses self-renewal, probably by destabilizing the messenger RNAs of Sox2 and/or Nanog (not Tbx3 or c-Myc as suggested by Buckley et al.[2]). In this model, changes in the cellular environment that counter this function of REST or stimulate the mRNA levels of Sox2 or Nanog could minimize the effect of REST. For this reason, in our study we used mESCs with a low passage number, and we cultured them without feeder cells to avoid possible contributions of the feeder cells or an adaptive response to high passage.

One possible reason for the discrepancies between our study and the studies by Jørgensen et al.[1] is the use of feeder cells in most of their experiments (see Fig. 1d of Jørgensen et al.). Indeed, our additional experiments showed that culturing mESCs with feeder cells severely dampens the requirement of REST-mediated maintenance of mESC self renewal and pluripotency (S.K.S., manuscript in preparation). However, to rule out adaptive responses, Jørgensen et al.[1] also performed REST knockdown experiments in feeder-independent "wild-type" mESCs (46C and OS25) using either si*Rest* or sh*Rest* (Fig. 1e of Jørgensen et al.[1]). However, 46C and OS25 are not wild-type cells and were generated from the E14TG2a wild-type cells by genetic manipulation (46C has an insertion in *Sox1* gene and OS25 has an insertion in *Sox2* and *Oct4* genes)[7]. It is unclear how these manipulations affected REST function and the use of these cells and not the wild-type cells probably explains their contradictory results (we used wild-type E14TG2a mESCs). In some experiments, Jørgensen et al.[1] used retinoic acid (see Fig. 1e of Jørgensen et al.). Additionally, we did not use retinoic acid in our experiments. Buckley et al.[2] mentioned that their data (mostly relative RT–PCR) are derived using feeder-free conditions. One potential reason for the discrepancy between our data and theirs could be high cell passage number or different cell type, among other experimental variations. For instance, their earlier report[6] mentioned above, in which they did find a critical role of REST in mESC self-renewal, used E14 ESCs instead of the HM1 cells used here. The absence of relative difference in *Mash1* and *Neurog1* levels in wild-type versus *Rest*[+/−] mESCs is not surprising because neuronal differentiation will cause a notable reduction in REST protein levels[8,9] (and consequent high *Mash1* and *Neurog1* expression) in both wild-type and *Rest*[+/−] cells. Our REST-miR-21 ChIP and functional data are specific, reproducible and significant. Buckley et al.[2] cited their high-throughput ChIP-PET and ChIP-qPCR analysis using our published primers[3] to indicate that REST does not bind to the miR-21 gene chromatin in mESCs. We published multiple locus-specific primer sets only for conventional ChIP but not ChIP-qPCR analysis (although these were available on request). The primers used in our ChIP-qPCR studies would provide good controls for binding[3], because binding of REST is weaker on miR-21 than on miR-124 chromatin and requires higher substrate. REST-miR-21 functional assays will also help resolve the discrepancies.

Buckley et al.[2] cited reports that used cell types other than mESCs. Moreover, their cited references 8 and 9 measure neither REST protein levels nor the direct effect of REST on miR-21. In contrast to Buckley et al.'s cited ref. 7, we note that our ref. 10 found that miR-21 expression was higher in differentiated than in undifferentiated mESCs, supporting our conclusions. Both communications cited Loh et al.[11] to suggest that knockdown of *Rest* using si*Rest* did not reduce self-renewal and did not induce loss of self-renewal markers in mESCs. This paper neither measured percentage self-renewal nor the level of the self-renewal regulators after si*Rest* treatment and cannot be used to counter our conclusions. Both communications also cited Chen et al.[12] to indicate that *Rest*[+/−] or *Rest*[−/−] mutant mice show germline transmission/gastrulation. This situation is similar to many other self-renewal regulators, such as LIF, LIF-receptor β, gp130 (also known as Il6st), Stat3 and c-Myc[13] and could be relevant during diapause[14] (see Supplementary Information of our paper for details).

Sanjay K. Singh[1,5], **Mohamedi N. Kagalwala**[1,5]†, **Jan Parker-Thornburg**[2], **Henry Adams**[1] & **Sadhan Majumder**[1,3,4,5,6]

[1]Department of Genetics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA.
e-mail: smajumder@mdanderson.org
[2]Department of Biochemistry and Molecular Biology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA.
[3]Department of Neuro-Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA.
[4]The Brain Tumor Center, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA.
[5]Center for Stem Cell and Developmental Biology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA.
[6]Program in Genes and Development, The University of Texas Graduate School of Biomedical Sciences at Houston, Houston, Texas 77030, USA.
†Present address: Laboratory of Genetics, Salk Institute for Biological Studies, La Jolla, California 92037, USA.

1. Jørgensen, H. F., Chen, Z.-F., Merkenschlager, M. & Fisher, A. G. Is REST required for ESC pluripotency? *Nature* **457,** doi:10.1038/nature07783 (2009).
2. Buckley, N. J., Johnson, R., Sun, Y.-M. & Stanton, L. W. Is REST a regulator of pluripotency? *Nature* **457,** doi:10.1038/nature07784 (2009).
3. Singh, S. K., Kagalwala, M. N., Parker-Thornburg, J., Adams, H. & Majumder, S. REST maintains self-renewal and pluripotency of embryonic stem cells. *Nature* **453,** 223–227 (2008).
4. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132,** 1049–1061 (2008).
5. Canzonetta, C. et al. DYRK1A-dosage imbalance perturbs NRSF/REST levels, deregulating pluripotency and embryonic stem cell fate in Down syndrome. *Am. J. Hum. Genet.* **83,** 388–400 (2008).
6. Johnson, R. et al. REST regulates distinct transcriptional networks in embryonic and neural stem cells. *PLoS Biol.* **6,** e256 (2008).
7. Ying, Q. L. & Smith, A. G. Defined conditions for neural commitment and differentiation. *Methods Enzymol.* **365,** 327–341 (2003).
8. Ballas, N., Grunseich, C., Lu, D. D., Speh, J. C. & Mandel, G. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell* **121,** 645–657 (2005).
9. Westbrook, T. F. et al. SCFbeta-TRCP controls oncogenic transformation and neural differentiation through REST degradation. *Nature* **452,** 370–374 (2008).
10. Houbaviy, H. B., Murray, M. F. & Sharp, P. A. Embryonic stem cell-specific microRNAs. *Dev. Cell* **5,** 351–358 (2003).
11. Loh, Y. H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38,** 431–440 (2006).
12. Chen, Z. F., Paquette, A. J. & Anderson, D. J. NRSF/REST is required *in vivo* for repression of multiple neuronal target genes during embryogenesis. *Nature Genet.* **20,** 136–142 (1998).
13. Davis, A. C. et al. A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. *Genes Dev.* **7,** 671–682 (1993).
14. Nichols, J., Chambers, I., Taga, T. & Smith, A. Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. *Development* **128,** 2333–2339 (2001).

CORRIGENDUM

Ecologically implausible carbon response?

Wim de Vries, Svein Solberg, Matthias Dobbertin, Hubert Sterba, Daniel Laubhann, Gert Jan Reinds, Gert-Jan Nabuurs, Per Gundersen & Mark A. Sutton

Nature **451**, E1–E3 (2008)

In this Brief Communication Arising, the name of Daniel Laubhann was incorrectly listed as Daniel Laubhahn.

Interhemispheric Atlantic seesaw response during the last deglaciation

Stephen Barker[1], Paula Diz[1]†, Maryline J. Vautravers[2], Jennifer Pike[1], Gregor Knorr[1]†, Ian R. Hall[1]
& Wallace S. Broecker[3]

The asynchronous relationship between millennial-scale temperature changes over Greenland and Antarctica during the last glacial period has led to the notion of a bipolar seesaw which acts to redistribute heat depending on the state of meridional overturning circulation within the Atlantic Ocean. Here we present new records from the South Atlantic that show rapid changes during the last deglaciation that were instantaneous (within dating uncertainty) and of opposite sign to those observed in the North Atlantic. Our results demonstrate a direct link between the abrupt changes associated with variations in the Atlantic meridional overturning circulation and the more gradual adjustments characteristic of the Southern Ocean. These results emphasize the importance of the Southern Ocean for the development and transmission of millennial-scale climate variability and highlight its role in deglacial climate change and the associated rise in atmospheric carbon dioxide.

The last glacial and deglacial periods were characterized by millennial-scale shifts in global climate. Records from Greenland ice cores[1] and North Atlantic sediments[2,3] suggest that high latitudes in the Northern Hemisphere were repeatedly subjected to large and abrupt fluctuations in temperature, commonly referred to as Dansgaard–Oeschger oscillations. In contrast to the abrupt changes observed in the north, temperature fluctuations over Antarctica were more gradual (warming and cooling over hundreds to thousands of years) and approximately out of phase with their northern counterparts[4]. Recent ice-core evidence from the Atlantic sector of Antarctica reveals a direct relationship between the extent of warming across Antarctica and the duration of cold, stadial conditions over Greenland[5]. The contrasting behaviour of temperature variations over Greenland and the North Atlantic as compared with Southern Hemisphere records has led to the notion of a bipolar seesaw, whereby changes in the strength of the Atlantic's conveyor circulation or, more precisely, the Atlantic meridional overturning circulation (AMOC) affect the distribution of heat between the South Atlantic and the North Atlantic and more widely[6,7]. Modelling results[8] suggest that a reduction in the strength of the AMOC (for example as may be caused by an input of fresh water to the North Atlantic) would give rise to an immediate decrease in northward heat transport. As a result, surface air temperatures over the North Atlantic would cool and those over the South Atlantic would warm. The transmission of these anomalies between the North Atlantic and the South Atlantic may take on the order of a few decades[8–10].

Although the bipolar seesaw can explain the antiphase relationship between Greenland and Antarctic temperature fluctuations through changes in the strength of the AMOC, it does not immediately account for the contrasting nature of the two signals. The simplest mechanism for transposing between the abrupt nature of Dansgaard–Oeschger oscillations and the more gradual temperature fluctuations observed over Antarctica involves the heat capacity of the Southern Ocean. This conceptual model (the "thermal bipolar seesaw"[11]) implies that temperature fluctuations in the South Atlantic should display an inverse form of the northern temperature

signal whereas those further south should be more gradual in nature[11]. However, so far all marine temperature reconstructions from the Southern Hemisphere (outside the tropics[12]) tend to resemble those from Antarctica (that is, with more gradual warming and cooling phases), and although a recent reconstruction from the southeast Pacific indicates a slightly more rapid response[13], a truly abrupt counterpart to the Northern Hemisphere variations has not yet been observed.

The last deglacial period (Termination I) witnessed several abrupt climatic shifts. During early Termination I ($\sim$18–14.6 kyr ago), temperature records from Greenland and the North Atlantic reveal a cold interval (Heinrich stadial 1 (HS1)) during which the massive ice-rafting episode known as Heinrich event 1 occurred (we use the term 'Heinrich stadial' to denote a stadial which contains a Heinrich event—a Heinrich stadial is not equivalent to a Heinrich event). This was followed by an abrupt and significant warming into the Bølling–Allerød warm interval before a rapid return to near-glacial conditions during the Younger Dryas event ($\sim$12.8–11.5 kyr ago). The final transition to Holocene warmth was completed by about 10 kyr ago. So far, changes observed in the Southern Hemisphere across Termination I were typically more gradual. Warming occurred during the period of severe North Atlantic cold (HS1) and was followed by a pause in warming (or a slight cooling), known as the Antarctic Cold Reversal, during the Bølling–Allerød interval. Completion of deglacial warming over Antarctica was accomplished during the Younger Dryas interval. A recent reconstruction of AMOC variability through Termination I[14] suggests that overturning within the Atlantic basin was severely affected during the cold episodes of Northern Hemisphere deglaciation. During HS1, cold conditions across the North Atlantic and warming over Antarctica were associated with a strong reduction in overturning strength[14]. A similar but less pronounced weakening occurred during the Younger Dryas. The slowdown of the AMOC and the coincident rise in Antarctic temperature and atmospheric CO_2 during the extreme (northern) cold of HS1 has been highlighted as a possible driving factor for the last termination[15]. This suggests that the bipolar seesaw

1097

is a key component not only for millennial-scale climate variability but also for changes on glacial–interglacial timescales. As such, increasing our understanding of the physical links between north and south at millennial timescales is critical for understanding the potential role of such variability as a feedback on global climate change[16].

Deglacial changes in the South Atlantic

Here we present continuous and high-resolution multi-proxy records spanning the late glacial and deglacial period from a marine sediment core (TNO57-21). Located in the southeast Atlantic Ocean (41.1° S, 7.8° E, 4,981-m water depth), at the northern margin of the modern Antarctic Circumpolar Current (ACC), the core is ideally positioned to capture an inverse form of the North Atlantic signal as predicted by the seesaw model (Fig. 1). Previous studies have illustrated the high quality of this core as a paleoceanographic archive (see, for example, refs 17, 18). We have measured 19 monospecific planktonic radiocarbon ages to provide an independent and absolute age model for the core (Supplementary Information). To investigate changes in surface ocean temperatures, we use a combination of planktonic foraminiferal assemblage counts and Mg/Ca ratios measured on the planktonic foraminifera *Globigerina bulloides*. We also analysed the benthic foraminiferal fauna and diatoms to investigate changes in the nature of productivity in the surface ocean.

Figure 2 shows our surface ocean records from TNO57-21 for the late glacial and deglacial period, plotted with temperature records from Greenland[1], Antarctica[5,19], the North Atlantic[2] and the southeast Pacific[13] along with the [231]Pa/[230]Th-based AMOC reconstruction of ref. 14. The relative abundance of polar planktonic foraminiferal species (*Neogloboquadrina pachyderma* (sinistral) and
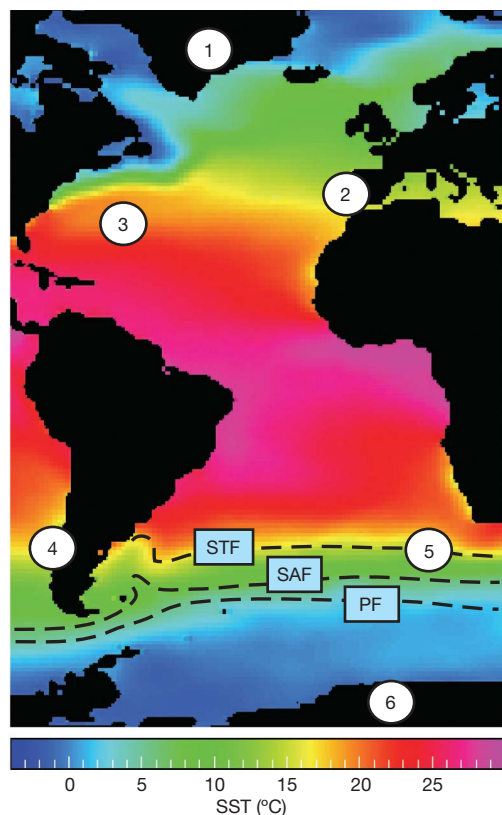
Figure 1 | Map showing sea surface temperature (SST)[50] and the positions of records shown in Fig. 2. Position 1, Greenland ice core GISP2 (ref. 1); 2, marine core SU8118 (ref. 2); 3, marine core OCE326-GGC5 (ref. 14); 4, marine core ODP 1233 (ref. 13); 5, marine core TNO57-21 (this study); 6, Antarctic ice core EDML (EPICA Dronning Maud Land)[5]. Dashed lines are approximate positions of the major oceanic fronts of the ACC: STF, Subtropical Front; SAF, Sub-Antarctic Front; PF, Polar Front.
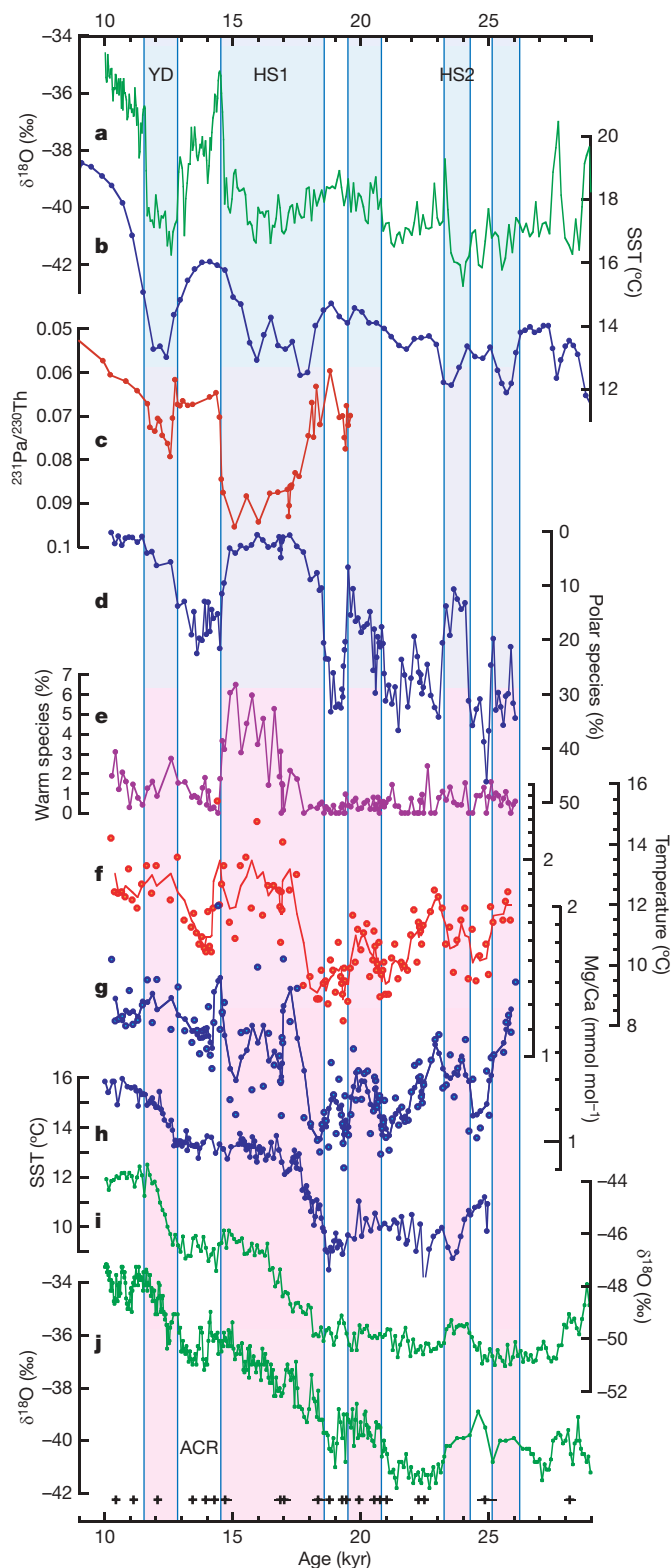
Figure 2 | Deglacial records from TNO57-21 plus other proxy records for temperature and circulation within the Atlantic Ocean. **a**, Greenland temperature (GISP2 ice core)[1]; **b**, North Atlantic SST (core SU8118)[2]; **c**, AMOC strength derived from sedimentary [231]Pa/[230]Th ratio[14] (higher values indicate a reduced AMOC); **d**, polar foraminiferal species in TNO57-21; **e**, warm species; **f**, Mg/Ca ratio (and calculated calcification temperature) in *G. bulloides*, adjusted to account for dissolution; **g**, measured Mg/Ca ratio in *G. bulloides* (solid lines in **f** and **g** are three-point running means); **h**, southeast Pacific SST[13]; **i, j**, Antarctic temperature (EDML[5] and Byrd[4,19] ice cores, respectively). Black symbols at the bottom indicate dating control points (calendar age uncertainty is 1σ; Supplementary Table 1). Shaded boxes represent periods of warmth or warming in the south according to our records.

Turborotalita quinqueloba; Fig. 2d) in TNO57-21 reveals large (>10%) and abrupt (~100-yr) changes throughout the interval. Moreover, within the uncertainties of our dating control (Supplementary Information) the abrupt shifts in assemblage are simultaneous with the abrupt transitions observed in the north. For example, the reduction in the AMOC (Fig. 2c) and associated cooling during HS1 (Fig. 2a, b) is matched by a sharp decline of polar species (a warming) in TNO57-21. A similar correspondence is observed during HS2 and the Younger Dryas. Conversely, warming events in the north (such as the Bølling–Allerød warming ~14.6 kyr ago and the Dansgaard–Oeschger event 2 ~23 kyr ago) are aligned with cooling in the south.

The abrupt nature of changes observed in TNO57-21 may be contrasted with the more gradual changes displayed in the temperature records from Antarctica (Fig. 2i, j) and the southeast Pacific (Fig. 2h). Periods of warming in the Antarctic record (for example during the Younger Dryas, HS1 and HS2) tend to begin in parallel with an abrupt decrease in polar species recorded in TNO57-21 and to end when polar species increase again (except for the end of the Younger Dryas, by which time mean conditions were presumably too warm for these species). Thus, the Antarctic signal resembles a damped, or integrated, form of the South Atlantic record. Our record also shows warming episodes ~21 and ~26 kyr ago. Both of these are aligned with warm intervals in the Byrd ice-core record from Antarctica[4] (Fig. 2j), and the earlier event apparently corresponds to a cool episode in the North Atlantic[2] (Fig. 2b). The warming ~21 kyr ago represents the end of significant ice-rafted debris at the site of TNO57-21[18] (Supplementary Fig. 8) and is apparent in records from both hemispheres (Fig. 2), perhaps representing an early phase of global deglacial warming.

The modern distribution of planktonic foraminiferal species in the South Atlantic and the adjacent sector of the Southern Ocean is dominated by the strong temperature gradients associated with the frontal systems of the ACC[20]. For example, polar species dominate the modern assemblage south of the present-day SAF but are virtually absent north of the STF. This change occurs within about 5° of latitude as a result of the SST gradient of >7 °C between the SAF and the STF (Fig. 1). Core TNO57-21 lies roughly beneath the STF in its modern position. As such, a relatively modest latitudinal shift or modulation of the frontal system (a few degrees) could cause a large (>20%) change in the abundance of polar species. We infer that the abrupt transitions in polar species observed in TNO57-21 during the deglacial period (Fig. 2d) represent shifts in the northern fronts associated with the ACC. Our results suggest a direct link between abrupt changes at high northern latitudes and similarly abrupt variations in the geometry of the ACC.

In the early part of the record, the abrupt decreases in polar species (notably HS2) were mainly accommodated by increases in the subpolar species, *N. pachyderma* (dextral). During deglaciation transitional species (for example *G. bulloides* and *Globorotalia inflata*) became more common and replaced both polar and subpolar species during the latest warming event. The appearance of warm-water species (*Globorotalia truncatulinoides* (dextral), *Globigerinoides ruber*, *Globorotalia hirsuta* and *Orbulina universa*; Fig. 2e) during HS1 (just before the Antarctic Cold Reversal) indicates that this was the warmest interval of the entire deglacial period, including the earliest Holocene, and highlights the anomalous conditions across the Atlantic during this time (in the modern South Atlantic, these species are only abundant well to the north of the STF[20]). It should be noted that we see no significant fluctuations in the contribution of faunas associated with the leakage of warm Indian Ocean water around the southern tip of Africa[21]. The 'Agulhas leakage' fauna in TNO57-21 show a relatively monotonic decrease throughout the interval, driven primarily by changes in the abundance of a single species, *Globigerinita glutinata*. We suggest that our record reflects a general warming of the South Atlantic during HS1 (as suggested by previous studies[22–24]) that may include the effect of increased Agulhas leakage during this time[21]. Furthermore, a reconstruction from 20° S in the western South Atlantic also suggests that sea surface temperatures in this region reached values comparable to, if not higher than, modern conditions during HS1[22].

Ambient temperature response

The more gradual appearance of warm-water foraminiferal species during HS1 (in contrast to the rapid fluctuations implied by the polar fauna) is reminiscent of Antarctic warming, that is, is influenced by the thermal inertia of a large water mass, and possibly reflects an increase in ambient temperatures in the South Atlantic region. To investigate the nature of such changes, we measured Mg/Ca ratios in the planktonic foraminifera *G. bulloides* to produce a species-specific temperature record (Fig. 2f, g). This species is common from the subtropics to the SAF[20]. Accordingly, its abundance in TNO57-21 is relatively less sensitive to the abrupt frontal changes at our site, possibly because it is able to alter its growth habitat (for example by altering its growth season and or depth habitat) to minimize more rapid environmental fluctuations. As such, it is potentially a good candidate for monitoring more gradual changes in the background or ambient conditions.

The record of Mg/Ca ratios measured in *G. bulloides* (Fig. 2g) displays millennial-scale variability of a different nature from the polar fauna record. According to the Mg/Ca ratios, warming generally occurs during intervals of reduced polar species (that is, 'warm' intervals) and vice versa. This is similar to the relationship between the polar species record and Antarctic temperature records. The Mg/Ca record displays a similar deglacial pattern to the SST record from the southeast Pacific[13] (Fig. 2h), but in contrast with that record the highest Mg/Ca ratios (warmest temperatures) in TNO57-21 are observed during HS1. The highest measured Mg/Ca ratios within HS1 occur near the beginning and right at the end of the interval, with lower values occurring in between (Fig. 2g). This pattern is unexpected, given the continuous increase in warm-water species throughout the same interval. TNO57-21 was recovered from a deep-water (4,981-m) location and has probably experienced partial dissolution. Dissolution is known to affect planktonic foraminiferal Mg/Ca ratios in a systematic way, with increasing dissolution causing a progressive decrease in Mg/Ca (ref. 25). The record of foraminiferal fragmentation from TNO57-21 suggests that dissolution has varied through time at this site (Supplementary Information). Specifically, the interval of low Mg/Ca ratios during HS1 is precisely aligned with an interval of increased fragmentation. We infer from this that the Mg/Ca ratios are lower than expected as a direct result of dissolution.

To generate a reasonable estimation of temperature from the measured record of Mg/Ca ratios, we have employed a straightforward algorithm for adjusting the Mg/Ca record, based on an empirical calibration of Mg/Ca versus foraminiferal fragmentation (see Supplementary Information for a discussion of this approach). The procedure restores the warmer temperatures during HS1 (Fig. 2f). In fact, these temperatures may reasonably have been even higher, depending on the choice of calibration employed. Of particular importance is the overall preservation of structure within the Mg/Ca record. This provides confidence that we are not introducing significant artefacts into the record through adjustment of the data. Furthermore, the adjusted Mg/Ca ratios yield an early-Holocene temperature consistent with the location of TNO57-21 (the calibration used for converting Mg/Ca ratios to temperature is taken from ref. 26). Our Mg/Ca results suggest significant variations in temperature (Fig. 2f). The rise in temperature during the early part of HS1 is 4–5 °C (with or without adjustment of the data). The decrease in temperature during the Antarctic Cold Reversal is 2–3 °C using a three-point running mean or 4–5 °C without averaging.

We note that planktonic foraminiferal assemblages may also be sensitive to dissolution in that increased fragmentation will result in a proportional increase of more robust species (which tend to represent colder temperatures[27]). However, because we observe a decrease

1099

in cold-water species (which are generally robust) and an increase in warm-water species (which are generally fragile) at the time of greatest fragmentation (Supplementary Information), we can be confident that these features are not an artefact of dissolution. In fact, the presence of warm-water species during HS1 may originally have been even greater than we observe.
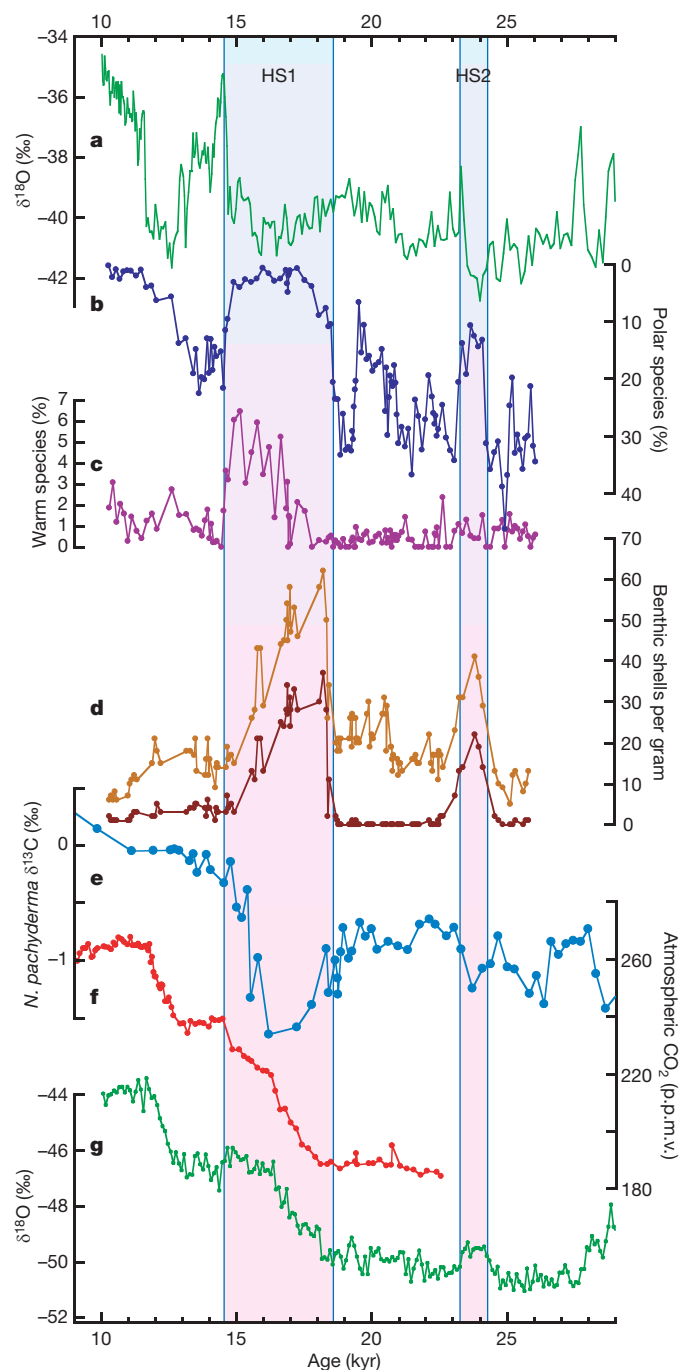

Figure 3 | Records of surface temperature and benthic fauna from TNO57-21. a, Greenland temperature (GISP2 ice core)[1]; **b**, polar foraminifera species in TNO57-21; **c**, warm foraminifera species; **d**, abundance of benthic foraminifera (upper curve) and *E. exigua* (lower curve)[17] ($\delta^{13}C = (^{13}C/^{12}C)_{sample}/(^{13}C/^{12}C)_{standard} - 1$; expressed relative to PeeDee Belemnite); **f**, atmospheric CO_2 (ref. 41; GICC05 timescale; see Supplementary Information); **g**, Antarctic temperature (EDML ice core)[5]. Shaded boxes represent Heinrich stadials and periods of Antarctic warming. Ice-core $\delta^{18}O$ values ($\delta^{18}O = (^{18}O/^{16}O)_{sample}/(^{18}O/^{16}O)_{standard} - 1$) are expressed relative to Vienna standard mean ocean water.

Surface ocean primary production

We also investigated changes in surface ocean productivity by analysing the benthic foraminiferal assemblages in TNO57-21 (Fig. 3d). The absolute abundance of benthic foraminifera shows two peaks (during HS1 and HS2). The magnitude of these peaks is driven largely by the presence of *Epistominella exigua*. In oligotrophic areas, this species is commonly associated with an increased flux of phytodetritus to the sea floor as a result of episodic phytoplankton blooms at the surface[28,29]. Today such events are commonly associated with the frontal systems of the Southern Ocean[30,31]. Along the STF, they occur as a direct consequence of turbulent eddy mixing across the front[31–33], which mixes nutrients from the sub-Antarctic zone into more stratified and warmer waters, thus creating suitable conditions for bloom development[34]. Increased surface production in the sub-Antarctic during the Heinrich events of Marine Isotope Stage (MIS) 3 has been reported previously[35], and it is significant that the episodes of increased *E. exigua* in TNO57-21 are associated only with the frontal movements inferred from the polar planktonic species during HS1 and HS2 and not with the shifts ~21 and ~26 kyr ago. This highlights the Heinrich stadials as unique within the deglacial sequence. We interpret the occurrence of *E. exigua* during HS1 and HS2 to reflect the proximity of the STF to TNO57-21 in combination with enhanced turbulence, possibly related to a strengthening of the ACC. This is supported by evidence from diatom assemblages in the same core (Supplementary Information). The sudden appearance of *E. exigua* at the beginning of HS1 is reminiscent of the abrupt change in polar species, and its more gradual decline coincides with the appearance of warm species and the increasing temperatures obtained from Mg/Ca ratios. Notably, a record of $\delta^{13}C$ in *N. pachyderma* (sinistral) from a nearby core[17] displays large negative values during the peaks of *E. exigua* abundance we observe in TNO57-21 (Fig. 3e). We suggest that these excursions reflect the advection of nutrient-rich waters across the STF during intervals of enhanced mixing along the front, supporting an increase in pulsed surface production.

Interhemispheric transmission

The characteristic responses of our proxy records may be placed in context by considering model simulations of abrupt changes in the AMOC (see, for example, refs 8–10, 36–38). Several mechanisms have been identified that might have an important role in the transmission of signals between the North Atlantic and the South Atlantic on timescales of decades[8–10,36] to several hundreds of years[9]. A persistent feature of the immediate (decadal) response to a slowdown in the AMOC is the southward shift of the intertropical convergence zone[8] and the westerly wind belts of the Southern Hemisphere[13,36]. A contemporaneous intensification of the ACC has also been reported as a direct consequence of increased westerly wind strength[10]. Our benthic fauna results suggest an increase in turbulent mixing across the northern fronts of the ACC during the AMOC slowdown of HS1, and therefore support the notion of an intensified ACC during this interval. Furthermore, the decline of polar foraminiferal species during Heinrich stadial events reflects the abrupt nature of southward shifts in the northern fronts of the ACC at these times. The 'instantaneous' bipolar mechanism is therefore not a straightforward temperature response, although some degree of rapid warming is predicted in the South Atlantic[8,9].

Our findings suggest that the ambient temperature response is more gradual, taking several hundreds of years. This reflects the build-up of heat within the South Atlantic as a consequence of a reduction in northward heat transport associated with a slowdown of the AMOC[6] combined with (during HS1) orbital forcing and possibly an increase in Agulhas leakage[21,37]. The propagation of thermal anomalies from the South Atlantic to Antarctica is slowed by the presence of the ACC[9], which represents a dynamical and thermal boundary between mid and high latitudes within the Southern Ocean[39] (note the delay between our ambient temperature increase

and Antarctic warming during HS1; Fig. 2f, i), although we would expect rapid transmission to other locations within the northern margin of the ACC (note the southeast Pacific record of ref. 13; Fig. 2f, h).

Our results appear to support a north-to-south transmission of abrupt changes across the Atlantic Ocean during deglaciation. However, they do not imply anything about the origin of such abrupt changes, which might be initiated by regional perturbations in the North Atlantic region (for example by freshwater forcing[8,10]) or as a nonlinear response to more gradual changes occurring remotely[40] (for example in the Southern Ocean)[37] or globally[38]. Furthermore, the Southern Ocean may provide a direct feedback on northern climate variability, for example through oceanic processes or atmospheric CO_2 (ref. 16).

CO_2 rise during Antarctic warming events

The close correspondence observed between atmospheric CO_2 and Antarctic temperature variability during the last deglaciation[41] (as well as during MIS 3[42]) indicates that the Southern Ocean has an important role in modulating atmospheric CO_2 on both glacial–interglacial and millennial timescales. Release of CO_2 during deglaciation may occur as a result of enhanced vertical mixing within the Southern Ocean[43,44], possibly as a result of a southerly shift in mid-latitude westerlies[43]. Intensification of the ACC during HS1, as suggested by our results and the modelling studies mentioned above, might have promoted the release of CO_2 by increasing the rate of vertical mixing within the Southern Ocean, thus contributing to the rise in atmospheric CO_2 at this time (Fig. 3f). This release of carbon may be reflected by the simultaneous 'leakage' of nutrients from within the Southern Ocean during HS1, as indicated by our benthic faunal records and the nearby record of planktonic $\delta^{13}C$ (ref. 17; Fig. 3). The more widespread occurrence of a planktonic $\delta^{13}C$ minimum event during the same interval[45] also supports this mechanism by the advection of nutrient- (and carbon-) rich waters away from the Southern Ocean, presumably by means of sub-Antarctic mode water formation[45].

The Heinrich stadials of MIS 3, and their corresponding Antarctic warming events, were also times of increasing atmospheric CO_2 (ref. 42), depleted planktonic $\delta^{13}C$ (ref. 17) and enhanced productivity in the sub-Antarctic region[35]. During the particularly prolonged interval of AMOC slowdown during HS1, the build-up of heat in the south and the release of CO_2 from the Southern Ocean (contributing to global warming in combination with, for example, insolation changes and ice-albedo feedbacks[46]) were apparently sufficient to enable deglaciation[15,38]. This mechanism may explain the rather abrupt onset of Antarctic temperature rise and the increase in atmospheric CO_2 at the start of deglaciation. We further speculate that during MIS 3 a similar mechanism was unable to drive deglaciation (possibly owing to orbital configuration[46] in combination with the subcritical size of continental ice sheets[47]), in which case the Antarctic warming events of MIS 3 may be thought of as 'failed terminations'.

METHODS SUMMARY

Sediment core samples were washed and sieved to 63 μm before drying and weighing. Planktonic foraminiferal faunal assemblage and fragment counts were performed on splits of the >150-μm fraction containing ~300 individual tests. Benthic foraminiferal counts were performed on the entire fraction >125 μm. Mg/Ca analyses were performed on typically 30 individual shells of *G. bulloides*, picked from the 250–315-μm fraction. Samples were prepared using the method of ref. 48 and analysed using a Finnigan Element XR high-resolution, inductively coupled plasma mass spectrometer. The age model for TNO57-21 was derived from 19 ^{14}C ages measured on *G. bulloides*. Two additional dates obtained from mixed species in a nearby core (RC11-83)[49] were used for the older part of the record. These were assigned depths on TNO57-21 by alignment of the records of percentage $CaCO_3$ from each core (Supplementary Fig. 1). ^{14}C ages were converted to calendar ages using a constant reservoir age of 600 yr.

1. Stuiver, M. & Grootes, P. M. GISP2 oxygen isotope ratios. *Quat. Res.* **53**, 277–283 (2000).
2. Bard, E., Rostek, F., Turon, J. L. & Gendreau, S. Hydrological impact of Heinrich events in the subtropical northeast Atlantic. *Science* **289**, 1321–1324 (2000).
3. Shackleton, N. J., Hall, M. A. & Vincent, E. Phase relationships between millennial-scale events 64,000-24,000 years ago. *Paleoceanography* **15**, 565–569 (2000).
4. Blunier, T. & Brook, E. J. Timing of millennial-scale climate change in Antarctica and Greenland during the last glacial period. *Science* **291**, 109–112 (2001).
5. EPICA Community Members. One-to-one coupling of glacial climate variability in Greenland and Antarctica. *Nature* **444**, 195–198 (2006).
6. Crowley, T. J. North Atlantic deep water cools the Southern Hemisphere. *Paleoceanography* **7**, 489–497 (1992).
7. Broecker, W. S. Paleocean circulation during the last deglaciation: A bipolar seesaw? *Paleoceanography* **13**, 119–121 (1998).
8. Vellinga, M. & Wood, R. A. Global climatic impacts of a collapse of the Atlantic thermohaline circulation. *Clim. Change* **54**, 251–267 (2002).
9. Schmittner, A., Saenko, O. A. & Weaver, A. J. Coupling of the hemispheres in observations and simulations of glacial climate change. *Quat. Sci. Rev.* **22**, 659–671 (2003).
10. Rind, D. *et al.* Effects of glacial meltwater in the GISS coupled atmosphere-ocean model - 2. A bipolar seesaw in Atlantic Deep Water production. *J. Geophys. Res.* **106**, 27355–27365 (2001).
11. Stocker, T. F. & Johnsen, S. J. A minimum thermodynamic model for the bipolar seesaw. *Paleoceanography* **18**, doi:10.1029/2003PA000920 (2003).
12. Wang, X. F. *et al.* Wet periods in northeastern Brazil over the past 210 kyr linked to distant climate anomalies. *Nature* **432**, 740–743 (2004).
13. Lamy, F. *et al.* Modulation of the bipolar seesaw in the southeast pacific during Termination 1. *Earth Planet. Sci. Lett.* **259**, 400–413 (2007).
14. McManus, J. F., Francois, R., Gherardi, J. M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428**, 834–837 (2004).
15. Denton, G. H., Broecker, W. S. & Alley, R. B. The mystery interval 17.5 to 14.5 kyrs ago. *PAGES News* **14**, 14–16 (2006).
16. Barker, S. & Knorr, G. Antarctic climate signature in the Greenland ice core record. *Proc. Natl Acad. Sci. USA* **104**, 17278–17282 (2007).
17. Ninnemann, U. S. & Charles, C. D. Regional differences in quaternary Subantarctic nutrient cycling: Link to intermediate and deep water ventilation. *Paleoceanography* **12**, 560–567 (1997).
18. Kanfoush, S. L. *et al.* Millennial-scale instability of the Antarctic ice sheet during the last glaciation. *Science* **288**, 1815–1818 (2000).
19. Johnsen, S. J., Dansgaard, W., Clausen, H. B. & Langway, C. C. Oxygen isotope profiles through Antarctic and Greenland ice sheets. *Nature* **235**, 429–434 (1972).
20. Niebler, H. S. & Gersonde, R. A planktic foraminiferal transfer function for the southern South Atlantic Ocean. *Mar. Micropaleontol.* **34**, 213–234 (1998).
21. Peeters, F. J. C. *et al.* Vigorous exchange between the Indian and Atlantic oceans at the end of the past five glacial periods. *Nature* **430**, 661–665 (2004).
22. Arz, H. W., Patzold, J. & Wefer, G. The deglacial history of the western tropical Atlantic as inferred from high resolution stable isotope records off northeastern Brazil. *Earth Planet. Sci. Lett.* **167**, 105–117 (1999).
23. Kim, J. H., Schneider, R. R., Muller, P. J. & Wefer, G. Interhemispheric comparison of deglacial sea-surface temperature patterns in Atlantic eastern boundary currents. *Earth Planet. Sci. Lett.* **194**, 383–393 (2002).
24. Farmer, E. C., deMenocal, P. B. & Marchitto, T. M. Holocene and deglacial ocean temperature variability in the Benguela upwelling region: Implications for low-latitude atmospheric circulation. *Paleoceanography* **20**, doi:10.1029/2004PA001049 (2005).
25. Barker, S., Cacho, I., Benway, H. & Tachikawa, K. Planktonic foraminiferal Mg/Ca as a proxy for past oceanic temperatures: a methodological overview and data compilation for the Last Glacial Maximum. *Quat. Sci. Rev.* **24**, 821–834 (2005).
26. Mashiotta, T. A., Lea, D. W. & Spero, H. J. Glacial-interglacial changes in Subantarctic sea surface temperature and $\delta^{18}O$-water using foraminiferal Mg. *Earth Planet. Sci. Lett.* **170**, 417–432 (1999).
27. Berger, W. H. Planktonic foraminifera: selective solution and the lysocline. *Mar. Geol.* **8**, 111–138 (1970).
28. Gooday, A. J. A response by benthic Foraminifera to the deposition of phytodetritus in the deep sea. *Nature* **332**, 70–73 (1988).
29. Cornelius, N. & Gooday, A. J. 'Live' (stained) deep-sea benthic foraminiferans in the western Weddell Sea: trends in abundance, diversity and taxonomic composition along a depth transect. *Deep-Sea Res. II* **51**, 1571–1602 (2004).
30. Froneman, P. W., Laubscher, R. K. & McQuaid, C. D. Size-fractionated primary production in the south Atlantic and Atlantic sectors of the Southern Ocean. *J. Plankton Res.* **23**, 611–622 (2001).
31. Saraceno, M., Provost, C. & Piola, A. R. On the relationship between satellite-retrieved surface temperature fronts and chlorophyll a in the western South Atlantic. *J. Geophys. Res.* **110**, doi:10.1029/2004JC002736 (2005).

32. Llido, J., Garcon, V., Lutjeharms, J. R. E. & Sudre, J. Event-scale blooms drive enhanced primary productivity at the Subtropical Convergence. *Geophys. Res. Lett.* **32,** doi:10.1029/2005GL022880 (2005).

33. Machu, E. *et al.* Phytoplankton distribution in the Agulhas system from a coupled physical-biological model. *Deep-Sea Res. I* **52,** 1300–1318 (2005).

34. Bathmann, U. V., Scharek, R., Klaas, C., Dubischar, C. D. & Smetacek, V. Spring development of phytoplankton biomass and composition in major water masses of the Atlantic sector of the Southern Ocean. *Deep-Sea Res. II* **44,** 51–67 (1997).

35. Sachs, J. P. & Anderson, R. F. Increased productivity in the subantarctic ocean during Heinrich events. *Nature* **434,** 1118–1121 (2005).

36. Timmermann, A., Krebs, U., Justino, F., Goosse, H. & Ivanochko, T. Mechanisms for millennial-scale global synchronization during the last glacial period. *Paleoceanography* **20,** doi:10.1029/2004PA001090 (2005).

37. Knorr, G. & Lohmann, G. Southern Ocean origin for the resumption of Atlantic thermohaline circulation during deglaciation. *Nature* **424,** 532–536 (2003).

38. Knorr, G. & Lohmann, G. Rapid transitions in the Atlantic thermohaline circulation triggered by global warming and meltwater during the last deglaciation. *Geochem. Geophys. Geosyst.* **8,** doi:10.1029/2007GC001604 (2007).

39. Cox, M. D. An idealized model of the world ocean. 1. The global-scale water masses. *J. Phys. Oceanogr.* **19,** 1730–1752 (1989).

40. Li, C., Battisti, D. S., Schrag, D. P. & Tziperman, E. Abrupt climate shifts in Greenland due to displacements of the sea ice edge. *Geophys. Res. Lett.* **32,** doi:10.1029/2005GL023492 (2005).

41. Monnin, E. *et al.* Atmospheric CO_2 concentrations over the last glacial termination. *Science* **291,** 112–114 (2001).

42. Ahn, J. & Brook, E. J. Atmospheric CO_2 and climate from 65 to 30 ka B.P. *Geophys. Res. Lett.* **34,** doi:10.1029/2007GL029551 (2007).

43. Toggweiler, J. R., Russell, J. L. & Carson, S. R. Midlatitude westerlies, atmospheric CO2, and climate change during the ice ages. *Paleoceanography* **21,** doi:10.1029/2005PA001154 (2006).

44. Kohler, P., Fischer, H., Munhoven, G. & Zeebe, R. E. Quantitative interpretation of atmospheric carbon records over the last glacial termination. *Glob. Biogeochem. Cycles* **19,** doi:10.1029/2004GB002345 (2005).

45. Spero, H. J. & Lea, D. W. The cause of carbon isotope minimum events on glacial terminations. *Science* **296,** 522–525 (2002).

46. Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the Earth's orbit: Pacemaker of the ice ages. *Science* **194,** 1121–1132 (1976).

47. Raymo, M. E. The timing of major climate terminations. *Paleoceanography* **12,** 577–585 (1997).

48. Barker, S., Greaves, M. & Elderfield, H. A study of cleaning procedures used for foraminiferal Mg/Ca paleothermometry. *Geochem. Geophys. Geosyst.* **4,** doi:10.1029/2003GC000559 (2003).

49. Charles, C. D., Lynch-Stieglitz, J., Ninnemann, U. S. & Fairbanks, R. G. Climate connections between the hemispheres revealed by deep sea sediment core/ice core correlations. *Earth Planet. Sci. Lett.* **142,** 19–27 (1996).

50. Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P. & Garcia, H. E. *World Ocean Atlas 2005, Volume 1: Temperature* (ed. Levitus, S.) (NOAA Atlas NESDIS 61, US Govt Printing Office, 2006).

Author Contributions S.B. designed the research and performed foraminiferal Mg/Ca analyses, P.D. performed benthic foraminiferal counts and picked planktonic foraminifera for ^{14}C dating and Mg/Ca analyses, M.J.V. performed planktonic foraminiferal counts, J.P. performed diatom counts and G.K. helped with interpretation. All authors contributed towards preparing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.B. (barkers3@cf.ac.uk).

METHODS

Age control. Age control for TNO57-21 was determined from a combination of 19 monospecific ^{14}C dates on *G. bulloides* (measured at the National Ocean Sciences Accelerator Mass Spectrometry Facility) and additional ^{14}C ages from a nearby core (RC11-83)[49]. The two cores were spliced together by means of their records of percentage $CaCO_3$, which contain sufficient structure to allow robust tuning with an estimated error $< \pm 10$ cm or ~600 yr (Supplementary Fig. 1). To attain sufficient sample weight for ^{14}C dating, tests of *G. bulloides* were combined from five consecutive 1-cm intervals of core. Therefore, each ^{14}C age represents a 5-cm interval (~300 yr of sedimentation). Tests of *G. bulloides* were picked from the 250–315-μm fraction or >250 μm where sample size was small. [14]C ages were converted to calendar ages using CALIB[51] software (version 5.0.1) with the Marine04 calibration curve[52] and $\Delta R = 200$ yr (that is, reservoir age of 600 yr). For samples older than 24 kyr (^{14}C) the calibration software of ref. 53 was employed (Supplementary Table 1). A surface reservoir correction of 600 yr was chosen to account for the relatively high southerly latitude of the core site and the possibility that glacial reservoir ages may have been slightly higher than modern in this region[54]. The age control points for RC11-83 given in Supplementary Table 1 were used to recalculate the age model for the $\delta^{13}C$ record of *N. pachyderma* (sinistral)[17] shown in Fig. 3 of the main text.

Faunal assemblage counts. Planktonic foraminiferal counts were performed on 1-cm intervals of sediment with a sampling frequency of 2 cm, following the taxonomy of ref. 55. Each sample was split to provide >300 individuals in the >150-μm fraction (only two samples contained fewer than 100 individuals). Planktonic foraminiferal fragment counts were made at the same time on the same sample splits. The abundance of benthic foraminifera was generally low (~100 individuals on average); therefore, all individuals from the >125-μm fraction were counted. Sampling frequency was every 1 to 5 cm for the benthic counts.

Mg/Ca ratios in *G. bulloides*. Mg/Ca ratios in *G. bulloides* were measured every 2 cm. Tests of *G. bulloides* were picked from the 250–315-μm fraction and prepared for analysis following the procedure of ref. 48, with an additional centrifuge step after dissolution. Typically 30 tests were used per sample; fewer where samples were small. Analyses were performed on a Finnigan Element XR (extended dynamic range) inductively coupled plasma mass spectrometer. Typical within-run precision was $0.35 \pm 0.2\%$, with long-term precision <2%. Simultaneous determination of Fe and Al allowed for the assessment of silicate contamination. One sample from a total of 124 was rejected on the basis of its Fe/Mg ratio[48] (Supplementary Fig. 2).

Atmospheric CO_2. The record of atmospheric CO_2 from ref. 41 was placed on the GICC05 timescale[56] (as used for the EDML $\delta^{18}O$ record[5]) by shifting the latest EDC3[57] timescale by 500 yr to align the records of CH_4 (Supplementary Fig. 9).

51. Stuiver, M. & Reimer, P. J. Extended ^{14}C data-base and revised Calib 3.0 ^{14}C age calibration program. *Radiocarbon* **35**, 215–230 (1993).
52. Hughen, K. A. *et al.* Marine04 marine radiocarbon age calibration, 0-26 cal kyr BP. *Radiocarbon* **46**, 1059–1086 (2004).
53. Fairbanks, R. G. *et al.* Radiocarbon calibration curve spanning 0 to 50,000 years BP based on paired Th-230/U-234/U-238 and C-14 dates on pristine corals. *Quat. Sci. Rev.* **24**, 1781–1796 (2005).
54. Bard, E. Correction of accelerator mass spectrometry ^{14}C ages measured in planktonic foraminifera: paleoceanographic implications. *Paleoceanography* **3**, 635–645 (1988).
55. Kennett, J. P. & Srinivasan, M. S. *Neogene Planktonic Foraminifera: A Phylogenetic Atlas* (Wiley, 1983).
56. Andersen, K. K. *et al.* A 60 000 year Greenland stratigraphic ice core chronology. *Clim. Past Discuss.* **3**, 1235–1260 (2007).
57. Parrenin, F. *et al.* The EDC3 chronology for the EPICA dome C ice core. *Clim. Past* **3**, 485–497 (2007).

ARTICLES

A mechanosensitive transcriptional mechanism that controls angiogenesis

Akiko Mammoto[1], Kip M. Connor[2], Tadanori Mammoto[1], Chong Wing Yung[1], Dongeun Huh[1], Christopher M. Aderman[2], Gustavo Mostoslavsky[3]†, Lois E. H. Smith[2] & Donald E. Ingber[1,4]

Angiogenesis is controlled by physical interactions between cells and extracellular matrix as well as soluble angiogenic factors, such as VEGF. However, the mechanism by which mechanical signals integrate with other microenvironmental cues to regulate neovascularization remains unknown. Here we show that the Rho inhibitor, p190RhoGAP (also known as GRLF1), controls capillary network formation *in vitro* in human microvascular endothelial cells and retinal angiogenesis *in vivo* by modulating the balance of activities between two antagonistic transcription factors, TFII-I (also known as GTF2I) and GATA2, that govern gene expression of the VEGF receptor *VEGFR2* (also known as *KDR*). Moreover, this new angiogenesis signalling pathway is sensitive to extracellular matrix elasticity as well as soluble VEGF. This is, to our knowledge, the first known functional cross-antagonism between transcription factors that controls tissue morphogenesis, and that responds to both mechanical and chemical cues.

Deregulation of angiogenesis—the growth of blood capillaries—contributes to the development of many diseases, including cancer, arthritis and blindness[1,2]. Most US Food and Drug Administration (FDA)-approved angiogenesis inhibitors target the oxygen-sensitive vascular endothelial growth factor, VEGF; however, neovascularization is also controlled by other microenvironmental signals, including mechanical forces conveyed by extracellular matrix (ECM). For example, although capillary development is driven by angiogenic mitogens, cell sensitivity to these soluble cues can be modulated by physical interactions between cells and ECM that alter cell shape and cytoskeletal structure[3–8]. Similar changes in capillary cell shape and function can be produced by changing ECM elasticity, adhesivity or topography, applying mechanical stresses, or altering cell-generated traction forces[3–8]. Mechanical tension also stimulates capillary growth and vascular remodelling *in vivo*[9], and regional variations of ECM mechanics and cell shape seem to mediate how neighbouring cells undergo localized differentials of growth and differentiation that drive three-dimensional tissue pattern formation[10,11]. However, the mechanism by which mechanical signals conveyed by ECM converge with those elicited by growth factors to control gene transcription required for angiogenic control remains unknown.

VEGF and its receptor VEGFR2 are of particular importance in angiogenesis because they are essential for normal blood vessel development[12,13], and deregulation of these factors leads to various pathological conditions[1,2,14]. We therefore investigated the role of VEGF and VEGFR2 in the mechanism by which these mechanical forces regulate capillary development. Analysis of cellular mechanotransduction during angiogenesis has shown that the small GTPase Rho mediates growth control *in vitro*[10,15], as well as blood vessel development *in vivo*[9], by modulating the mechanical force balance that governs cell shape. Stress-induced distortion of the capillary cell cytoskeleton regulates Rho activity by controlling its upstream inhibitor, p190RhoGAP[16]. p190RhoGAP was shown to bind to the transcription factor TFII-I and sequester it in the cytoplasm of fibroblasts[17], and

TFII-I is a multifunctional transcription factor that regulates *VEGFR2* expression in large vessel endothelial cells[18,19]. Although it remains unknown whether it has a role in angiogenesis, TFII-I deletions are associated with cardiovascular defects[20]. These findings raise the possibility that TFII-I might also contribute to intracellular signalling mechanisms triggered by mechanical forces during vascular network formation.

p190RhoGAP and TFII-I regulate VEGFR2

To explore whether p190RhoGAP modulates vascular development by altering gene transcription, we knocked down *p190RhoGAP* in human microvascular endothelial (HMVE) cells using short-interfering RNA (siRNA). TFII-I protein levels were approximately 1.5 times higher in the nuclear fraction of knockdown cells (Supplementary Fig. 2a), and this was confirmed by fluorescence microscopy (Supplementary Fig. 2b). As first shown in fibroblasts[17], we found that p190RhoGAP co-precipitates with TFII-I (and vice versa) (Supplementary Fig. 2c), suggesting that p190RhoGAP may also bind TFII-I and sequester it in the cytoplasm of human capillary cells.

TFII-I upregulates VEGFR2 protein expression in human aortic endothelial cells by binding to the *VEGFR2* promoter[18]. However, *TFII-I* knockdown in our HMVE cells increased (rather than decreased) *VEGFR2* messenger RNA and protein levels by two- to threefold relative to control cells, and overexpressing *TFII-I* (delta isoform[21]) using lentiviral transduction produced the opposite effect (Fig. 1a and Supplementary Fig. 3a). This may be due to the fact that macrovascular and microvascular endothelial cells undergo distinct morphogenetic programs (produce large tubes versus branching capillaries).

TFII-I binds to the initiator (Inr) region of the *VEGFR2* gene promoter, and smaller portions of this promoter (human *VEGFR2*: −225 to +268, −570 to +268, and −780 to +268) have similar or greater activity compared to the full-length promoter (4 kilobases (kb) long)[18,22]. When these various *VEGFR2* promoters were characterized

[1]Vascular Biology Program, Departments of Pathology & Surgery, [2]Department of Ophthalmology, Children's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Department of Genetics, Harvard Medical School, Harvard Institute of Medicine, Boston, Massachusetts 02215, USA. [4]Wyss Institute for Biologically Inspired Engineering and Harvard School of Engineering and Applied Sciences, Cambridge, Massachusetts 02138, USA. †Present address: Department of Medicine, Boston University School of Medicine, Boston, Massachusetts 02118, USA.
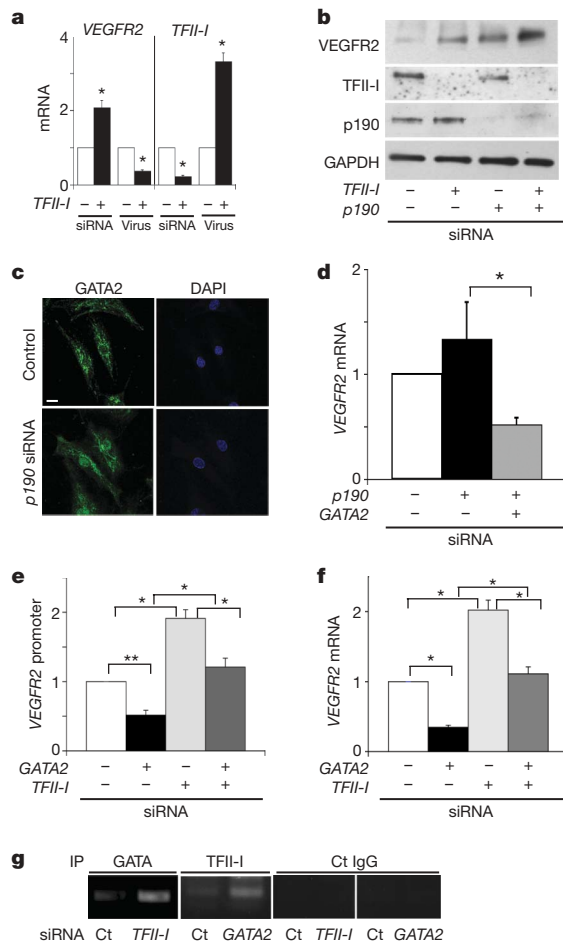
Figure 1 | TFII-I and GATA2 control *VEGFR2* expression via p190RhoGAP.
a, *VEGFR2* and *TFII-I* mRNA levels in cells treated with *TFII-I* siRNA or
lentiviral vectors (virus) relative to control cells (*$P < 0.05$; unpaired
Student's *t*-test is used throughout). **b**, Immunoblots showing VEGFR2,
TFII-I, p190RhoGAP (p190) and GAPDH protein levels in cells treated with
TFII-I or *p190RhoGAP* siRNAs, or both. **c**, Immunofluorescence
micrographs showing GATA2 distribution and nuclear 4,6-diamidino-2-
phenylindole (DAPI) staining in control and *p190RhoGAP* knockdown cells
cultured with 0.3% serum (scale bar, 5 μm). **d**, *VEGFR2* mRNA levels in cells
transfected with *p190RhoGAP* siRNA alone or with *GATA2* siRNA
(*$P < 0.01$). **e, f**, *VEGFR2* promoter activities (**e**) and mRNA levels (**f**) in
cells transfected with *GATA2* or *TFII-I* siRNA alone or in combination
(*$P < 0.01$, **$P < 0.05$). **g**, ChIP analysis showing *VEGFR2* promoter co-
immunoprecipitating (IP) with GATA2 or TFII-I antibodies in cells
transfected with *TFII-I* or *GATA2* siRNA ($n = 3$). Ct, control. All error
bars are s.e.m.

using a luciferase assay in human umbilical vein endothelial (HUVE)
cells, *TFII-I* knockdown increased, and overexpression of *TFII-I* using
DNA transfection decreased, *VEGFR2* promoter activity by 1.5 to
twofold and one-fifth normal levels, respectively (Supplementary
Fig. 3b). The promoter-less pGL3 basic reporter showed no promoter
activity (data not shown). Moreover, mutagenesis of the *VEGFR2* Inr
(−780 to +268; 780MUT) decreased promoter activity in both con-
trol and *TFII-I* knockdown cells (Supplementary Fig. 3b), confirming
that TFII-I decreases *VEGFR2* promoter activity through the Inr
region. Importantly, the specificity of these effects of *TFII-I* knock-
down was confirmed by demonstrating that reconstitution of mouse
TfII-I can reverse these effects (Supplementary Fig. 3c).

We next addressed whether p190RhoGAP stimulates *VEGFR2*
expression by restricting the nuclear translocation of TFII-I
(Supplementary Fig. 2a, b). *TFII-I* knockdown increased, and
p190RhoGAP knockdown decreased, *VEGFR2* promoter activity
(−780 to +268) relative to control cells (Supplementary Fig. 3d).

The double knockdown of *p190RhoGAP* and *TFII-I* exhibited the
same stimulation as observed in the *TFII-I* knockdown cells
(Supplementary Fig. 3d). However, *p190RhoGAP* knockdown did
not decrease *VEGFR2* mRNA and protein levels; in fact, it increased
VEGFR2 protein levels (Fig. 1b and Supplementary Fig. 3d).
Moreover, *p190RhoGAP* and *TFII-I* double knockdown produced
similar or slightly higher levels of *VEGFR2* mRNA and protein
expression (Fig. 1b and Supplementary Fig. 3d) compared to either
p190RhoGAP or *TFII-I* knockdown alone. These results raised the
intriguing possibility that an antagonist of TFII-I activity might exist
that also contributes to p190RhoGAP-dependent control of *VEGFR2*
expression.

GATA2 upregulates VEGFR2

GATA2 is another transcription factor that binds to the *VEGFR2*
promoter and regulates its activity[23,24]. Because many GATA family
members antagonize the effects of other transcription factors at
promoter sites[25–28], we explored whether GATA2 mediates p190RhoGAP-
dependent control of *VEGFR2* expression by opposing TFII-I activity.
When *p190RhoGAP* was knocked down, GATA2 levels in the nucleus
increased markedly (>10-fold) relative to control cells even though
total GATA2 levels remained the same, as shown by immunoblotting
(Supplementary Fig. 3e) and fluorescence microscopy (Fig. 1c).
p190RhoGAP also co-immunoprecipitated with GATA2 and vice
versa (Supplementary Fig. 3f). Notably, this increase in nuclear
GATA2 was significantly higher than the 1.5-fold increase in nuclear
TFII-I produced using *p190RhoGAP* siRNA (Supplementary Fig. 2a,
b), suggesting that p190RhoGAP binds and sequesters GATA2 in the
cytoplasm more efficiently than TFII-I in capillary cells. *GATA2*
knockdown using siRNA decreased *VEGFR2* promoter activity, as
well mRNA and protein levels (Supplementary Fig. 4a, b) as previously
observed[23], whereas *GATA2* overexpression produced the opposite
effect. These effects of *GATA2* siRNA on *VEGFR2* promoter activity
and protein expression were specific as they were reversed by recon-
stitution of the mouse *Gata2* gene (Supplementary Fig. 3c).
p190RhoGAP knockdown increased the expression of *VEGFR2*
mRNA and protein, and double knockdown with *GATA2* inhibited
these effects (Fig. 1d and Supplementary Fig. 4c). Thus, GATA2 seems
to upregulate *VEGFR2* promoter activity and mediate the effects of
p190RhoGAP on *VEGFR2* gene expression in capillary cells. However,
knockdown of *p190RhoGAP*, which releases more GATA2 than TFII-I
and increases *VEGFR2* expression, did not increase *VEGFR2* promoter
activity (Supplementary Fig. 3d). This may be because cellular *TFII-I*
levels are five times higher than *GATA2* levels (data not shown). We
also measured *VEGFR2* promoter activity using only a portion of its
promoter, and hence, TFII-I and GATA2 may also exert regulatory
activities at other promoter sites; alternatively, p190RhoGAP might
elicit signals that influence mRNA stability.

We next determined whether GATA2 and TFII-I directly antagonize
each other. Simultaneous knockdown of *TFII-I* and *GATA2* abrogated
each other's effects on *VEGFR2* promoter activity (Fig. 1e), mRNA
production (Fig. 1f) and protein expression levels (Supplementary
Fig. 4e), and simultaneous overexpression of *TFII-I* and *GATA2*
produced similar effects (Supplementary Fig. 4d, e). These effects were
specific for *VEGFR2* as the knockdown of *TFII-I* or *GATA2* did not
alter the expression of *VEGFR1* (also known as *FLT1*) or *VEGFR3*
(*FLT4*) in HMVE cells (Supplementary Fig. 5a). Notably, although
p190RhoGAP is a Rho inhibitor, altering Rho activity with constitu-
tively active RHOA, membrane-permeable C3 exoenzyme or siRNA
directed to another Rho-inhibiting RhoGAP (*p73RhoGAP*, also known
as *ARHGAP24*)[29] did not change *VEGFR2* mRNA or protein levels
(Supplementary Fig. 5b, c). Thus, p190RhoGAP seems to control
VEGFR2 expression solely by its ability to sequester these transcription
factors.

Analysis of this mechanism of functional cross-antagonism demon-
strated that GATA2 and TFII-I associate with each other, as GATA2
co-immunoprecipitated with TFII-I, and vice versa (Supplementary

Fig. 6a). Furthermore, GATA2 even binds to TFII-I in *p190RhoGAP* knockdown cells and to p190RhoGAP in *TFII-I* knockdown cells (Supplementary Fig. 6b). These results indicate that the various heterodimeric combinations of TFII-I, GATA2 and p190RhoGAP exist in separate pools, and that these heterodimers then associate to form a larger ternary complex (Supplementary Fig. 1).

Chromatin immunoprecipitation (ChIP) analysis showed that *TFII-I* knockdown cells exhibited increased recruitment of GATA2 to the GATA binding site (-150 to $+150$) compared to control cells and vice versa (Fig. 1g), whereas control IgG did not immunoprecipitate these DNAs (Fig. 1g). *p190RhoGAP* knockdown decreased the recruitment of TFII-I, but not GATA2, to this promoter site, which resulted in a relative net increase in GATA2 recruitment to this site (Supplementary Fig. 6c). TFII-I and GATA2 therefore compete with each other for occupancy of a common region of the *VEGFR2* promoter, which is controlled by p190RhoGAP.

Mechanical control of VEGFR2

Soluble growth factors, integrin binding to ECM, and mechanical distortion of the cytoskeleton all regulate p190RhoGAP activity in cells[30]. Cell binding to growth factors and adhesive contact formation with ECM also control *VEGFR2* expression[31,32], and soluble mitogens (5% serum plus VEGF, bFGF and PDGF) increase nuclear translocation of TFII-I and GATA2 in HMVE cells (Supplementary Fig. 7a). We therefore next explored whether changes in mechanical interactions between cells and ECM regulate this pathway as well. When HMVE cells were cultured in the absence of mitogens on fibronectin-coated polyacrylamide gels with different elasticity (Young's moduli (stiffness) of 150 to 4,000 Pa), they appeared round on the soft gels, whereas they flattened on the stiffer gels, as previously observed[33,34], which is based on differences in their ability to physically resist cell traction forces[8,33,34]. Nuclear GATA2 levels were significantly higher in cells on the stiffer gels, whereas nuclei exhibit similar high levels of TFII-I regardless of ECM stiffness (Fig. 2a), and similar results were

obtained in the presence of several soluble factors or VEGF alone (Supplementary Fig. 7b). Moreover, *VEGFR2* mRNA and protein levels were higher in cells on the stiffer (4,000 Pa) gels (Fig. 2b and Supplementary Fig. 7c). Interestingly, this relatively stiff, but still compliant, ECM gel seemed to support optimal responsiveness of this signalling pathway because *VEGFR2* mRNA and protein levels were significantly lower in cells cultured on rigid glass ECM substrates (Fig. 2b, and Supplementary Fig. 7c). ECM elasticity may therefore regulate *VEGFR2* expression preferentially via GATA2, particularly over the stiffness range analysed in this study.

Further analysis showed that *TFII-I* knockdown restored *VEGFR2* expression in cells on soft gels to levels similar to those in cells on stiff gels, and double knockdown with *GATA2* (which decreases *VEGFR2* expression) abrogated the effects of *TFII-I* (Fig. 2c and Supplementary Fig. 7d). *p190RhoGAP* knockdown also increased the expression of *VEGFR2* mRNA and protein, and double knockdown with *GATA2* inhibited these effects on soft gels (Fig. 2d and Supplementary Fig. 7d). These data indicate that p190RhoGAP and the mutually antagonistic TFII-I and GATA2 transcription factors mediate the effects of ECM elasticity on *VEGFR2* expression in these cells, which is preferentially shifted to GATA2 on stiffer gels.

Transcriptional control of angiogenesis

We next explored the functional relevance of this antagonism between GATA2 and TFII-I by analysing capillary cell migration and differentiation (tube formation) *in vitro*. When HMVE cells were transfected with *TFII-I* siRNA and analysed using the transwell migration assay, VEGF-stimulated cell motility increased by twofold, whereas this induction was prevented by knocking down *GATA2* simultaneously with *TFII-I* (Fig. 3a). The VEGFR2 kinase inhibitor SU5416 totally inhibited the migration of these cells (Fig. 3a), confirming that these effects are mediated by VEGFR2 signalling. Moreover, *GATA2* overexpression increased VEGF-induced cell migration by threefold, and the simultaneous overexpression of *GATA2* and *TFII-I*
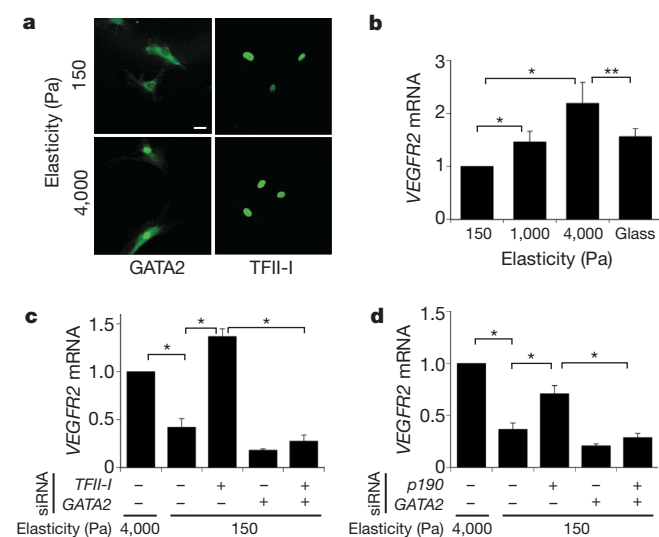
Figure 2 | Matrix elasticity controls *VEGFR2* expression via TFII-I and GATA2. a, Immunofluorescence micrographs showing GATA2 and TFII-I distribution in HMVE cells cultured on the fibronectin-coated gels of different stiffness (Young's moduli of 150 and 4,000 Pa) in EBM2 medium with 0.3% serum (scale bar, 5 μm). **b,** *VEGFR2* mRNA levels in HMVE cells cultured on rigid fibronectin-coated glass or the gels of different elasticity (150, 1,000 and 4,000 Pa; normalized to that in cells on the softest gels; *$P < 0.01$, **$P < 0.05$. **c,** *VEGFR2* mRNA levels in HMVE cells on soft gels (150 Pa) transfected with *TFII-I* or *GATA2* siRNA alone or in combination (normalized to cells on the stiffest gels; *$P < 0.01$). **d,** *VEGFR2* mRNA levels in HMVE cells on soft gels transfected with *p190RhoGAP* or *GATA2* siRNA alone or in combination (normalized to cells on stiffest gels; *$P < 0.01$). Error bars represent s.e.m. of three replica experiments.
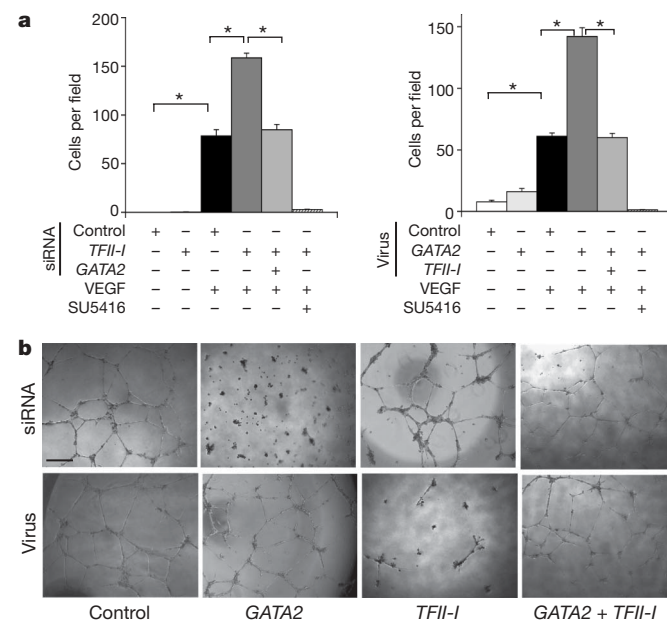
Figure 3 | Antagonism between GATA2 and TFII-I controls capillary cell migration and tube formation *in vitro*. a, The motility of HMVE cells transfected with human siRNAs or transduced with lentiviral vectors encoding *GATA2* or *TFII-I*, alone or in combination, was quantified using the transwell migration assay (*$P < 0.01$). Where indicated, VEGF (10 ng ml^{-1}) was added to the lower chamber and SU5416 was added in both chambers. Error bars represent s.e.m. of three replica experiments. **b,** Micrographs showing *in vitro* tube formation induced by VEGF (10 ng ml^{-1}) in HMVE cells transfected with siRNAs or transduced with lentiviral vectors encoding *GATA2* or *TFII-I*, alone or in combination (scale bar, 500 μm).
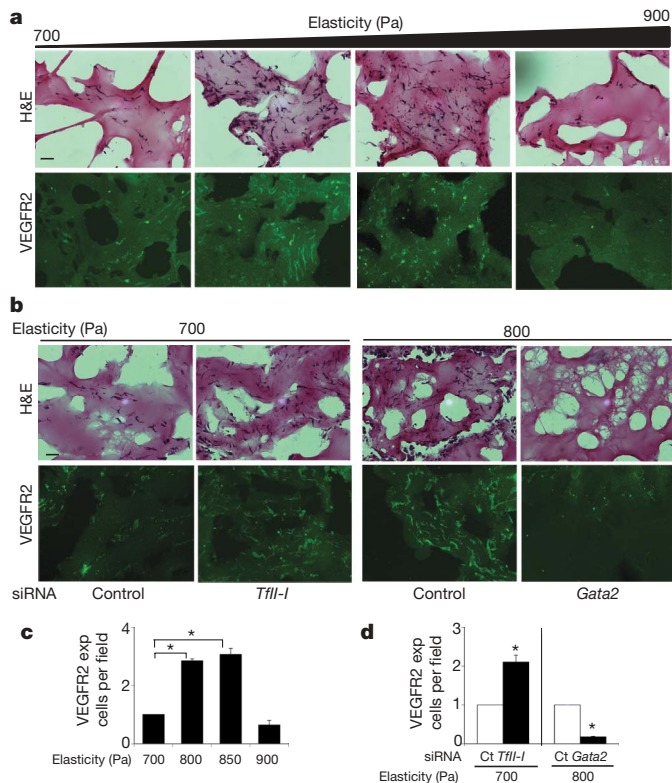
1105

Figure 4 | Matrix elasticity controls vessel formation via TFII-I and GATA2 *in vivo*. a–d, Light (haematoxylin and eosin (H&E)-stained) (**a, b**, top) or immunofluorescence micrographs (**a, b**, bottom), and vessel densities per high power field (**c, d**), showing cell infiltration and VEGFR2 expression (exp) in Matrigel with different elasticity implanted in mice for 7 days without (**a, c**) or with *TfII-I* or *Gata2* siRNAs (**b, d**). Scale bars, 25 μm. The number of VEGFR2-positive blood vessels was normalized to that in the 700 Pa gels in **c** and to that in gels treated with control siRNA in **d** (*n* = 6). Data are mean ± s.e.m., *P < 0.01.

abolished these effects (Fig. 3a). These effects were specific in that reconstitution of *TfII-I* or *Gata2* reversed the effects of knocking down *TFII-I* or *GATA2*, respectively (Supplementary Fig. 8a).

We then studied capillary tube formation by HMVE cells cultured within the ECM gel Matrigel, which supports angiogenesis in part because of its flexibility. Tube formation was stimulated by VEGF in a dose-dependent manner, whereas bFGF (also known as FGF2) and PDGFB were ineffective, and the effects of VEGF were inhibited by SU5416 (Supplementary Fig. 8b). *GATA2* knockdown and *TFII-I* overexpression suppressed VEGF-stimulated capillary development, and once again either knocking down or overexpressing both transcription factors simultaneously negated these effects (Fig. 3b and Supplementary Fig. 8c). *TFII-I* knockdown or *GATA2* overexpression alone did not produce significant effects on tube formation, apparently because it was already optimally stimulated under these conditions (Fig. 3b); knockdown or overexpression of *p190RhoGAP* also did not alter tube formation (Supplementary Fig. 8d). Thus, functional antagonism between GATA2 and TFII-I at the level of *VEGFR2* transcription translates into biologically relevant changes in capillary cell behaviour that are required for formation of three-dimensional capillary networks.

Control of angiogenesis *in vivo*

We next examined whether ECM mechanics govern vessel formation *in vivo* using a modified Matrigel implant assay. Maximal levels of cell infiltration, capillary blood vessel formation, and VEGFR2 expression were observed in Matrigel with intermediate stiffness (800 Pa) compared to cells in more or less rigid gels (900 or 700 Pa, respectively) (Fig. 4a, c and Supplementary Fig. 9a, b). VEGFR2, GATA2 and
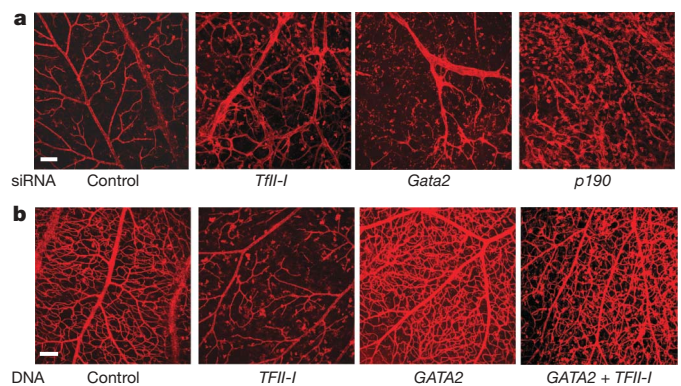
Figure 5 | TFII-I, GATA2 and p190RhoGAP regulate retinal vessel formation *in vivo*. a, Confocal fluorescence micrographs showing projected images of vessels stained with Alexa-594-isolectin in the retina of a control mouse eye versus eyes transfected with *TfII-I*, *Gata2* or *p190RhoGAP* siRNAs (scale bar, 0.1 mm). **b**, Confocal fluorescence micrographs showing projected images of vessels stained with Alexa-594-isolectin in the retina of a control mouse eye versus eyes with overexpressed human *TFII-I* or *GATA2*, alone or in combination (scale bar, 0.1 mm).

TFII-I all localized within cells lining CD31-positive staining micro-vessels (Supplementary Fig. 9c). The finding that the optimal ECM stiffness required for angiogenesis *in vivo* (800 Pa) was different from that observed *in vitro* (4,000 Pa) is probably due to different requirements by cells when cultured on two-dimensional ECM versus a three-dimensional ECM gel; it may also relate to how the Matrigel is remodelled over time *in vivo*.

We next performed siRNA-mediated gene knockdown in the *in vivo* Matrigel assay. Notably, *TfII-I* knockdown increased cell infiltration, capillary vessel formation and VEGFR2 expression by cells in the soft gels so that it mimicked the behaviour of cells on the intermediate stiffness gel, whereas *Gata2* knockdown produced the opposite effects (Fig. 4b, d and Supplementary Fig. 10a), and *p190RhoGAP* knockdown also increased the level of vessel formation in the soft gels (Supplementary Fig. 10b). These results indicate that TFII-I, GATA2 and p190RhoGAP mediate the signalling effects of ECM mechanics on vessel formation *in vivo*.

To unequivocally confirm the functional and clinical relevance of the regulatory interactions between TFII-I, GATA2 and p190RhoGAP, we modulated their expression in the neonatal mouse retina because angiogenesis in this growing organ is tightly regulated by VEGF and its receptors[14,35,36]. TFII-I, GATA2 and VEGFR2 localized to the three layers of the retina where blood vessels are located in postnatal day (P)15 mice (Supplementary Fig. 11a, b). Consistent with *in vitro* data, knocking down *TfII-I* using intravitreal injection of siRNA to P14 mice resulted in increased *Vegfr2* expression and the appearance of highly tortuous dilated vessels and a significant increase in vascular density in the retina, whereas *Gata2* knockdown suppressed *Vegfr2* expression, disrupted capillary network formation and decreased vascular density (Fig. 5a and Supplementary Fig. 12a, b). Knockdown of *p190RhoGAP* also increased *Vegfr2* expression and vessel density, but it resulted in a slightly different capillary growth pattern, perhaps in part because of its known effects on vascular permeability[37] (Fig. 5a and Supplementary Fig. 12a, b). Similar effects were observed in earlier (P5) retina that contain rapidly growing microvessels (Supplementary Fig. 13). Furthermore, overexpression of human *TFII-I* and *GATA2* produced opposite effects on vascular density in P14 retina, and simultaneous overexpression of both factors abolished these effects, confirming our *in vitro* findings (Fig. 5b and Supplementary Fig. 12c). Knockdown of *p190RhoGAP* or *TfII-I* did not significantly change *Vegf* expression levels in the retina, whereas *Gata2* knockdown increased (rather than decreased) its expression (Supplementary Fig. 12d). The decrease of *Vegfr2* expression seems to be sufficient to abrogate this *Vegf* response in *Gata2* knockdown retina, and thereby inhibit angiogenesis. Hence,

p190RhoGAP controls *Vegfr2* expression and vascular development by modulating the balance between TFII-I and GATA2 activities such that TFII-I activity dominates (and *Vegfr2* expression is suppressed) in the retina.

Discussion

Transcription factors change their activities in a spatio-temporal manner during development, and thereby specify cell fate[38,39]. Here we show that p190RhoGAP, which has been shown to be regulated by growth factors, ECM binding and cytoskeletal distortion, also controls *VEGFR2* expression, as well as angiogenesis *in vitro* and *in vivo*, by altering the balance between two mutually antagonistic transcription factors: TFII-I and GATA2 (Supplementary Fig. 1). Moreover, we demonstrate that p190RhoGAP and this downstream transcriptional control mechanism are regulated by mechanical signals conveyed by variations in ECM elasticity. This mechanism is analogous to other developmental mechanisms used by haematopoietic cells and mammary epithelium[28,38–40]; however, this is, to our knowledge, the first demonstration that transcriptional cross-antagonism can govern histodifferentiation and tissue morphogenesis, and be sensitive to mechanical as well as chemical cues.

Our data suggest that an appropriate level of ECM stiffness may be required for optimal *VEGFR2* expression and vascular development *in vitro* and *in vivo*. In fact, the fates of different cell types are exquisitely sensitive to distinct ECM elasticity values that often match those exhibited by their host tissues[41]. Abrupt local changes in ECM mechanics also accompany the switch between active growth and quiescent differentiation of functional capillary networks in living tissues[42]. Cell rounding suppressed p190RhoGAP activity within 1 to 2 h by altering its binding to the cytoskeletal protein filamin[16]. A similar cytoskeleton-based effect could mediate the effect of ECM elasticity on p190RhoGAP activity at later times, and thereby control the TFII-I:GATA2 balance and *VEGFR2* transcription. Because VEGFR2 is expressed in neurons as well as in endothelial cells in the retina, some of the effects of gene knockdown we observed may not be specific to capillary cells. However, even these alterations might be relevant for control of vascular development because neuron–vessel interactions are indispensable for normal microvasculature function and patterning[43].

These data demonstrate a previously unknown mechanosensitive signalling pathway that controls *VEGFR2* promoter activity and expression, and which represents a point of convergence for all three classes of microenvironmental signals that regulate capillary morphogenesis. Development of specific modifiers of this pathway could therefore lead to new therapeutic approaches for various angiogenesis-dependent diseases, including proliferative retinopathy, arthritis and cancer, in the future.

METHODS SUMMARY

Expression of *TFII-I*, *GATA2* and *VEGFR2* were evaluated by quantitative PCR with reverse transcription (qRT–PCR) and immunoblotting. A luciferase reporter assay was used to measure *VEGFR2* promoter activity. To test the effects of TFII-I and GATA2 on *VEGFR2* expression and angiogenesis *in vitro*, siRNA-mediated knockdown or lentiviral transduction was performed in HMVE cells. *In vitro* analysis of angiogenesis was carried out using transwell migration and Matrigel tube formation assays, and similar results were obtained using both native and growth factor-reduced forms of Matrigel. A subcutaneous Matrigel angiogenesis assay was used to analyse the effects of ECM mechanics on capillary formation *in vivo*. Retinal vessel formation was also studied in newborn C57BL/6 mice, and gene expression was manipulated in whole living retina by intravitreal injection of siRNA or DNA into the eye at P5 or P14.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Ferrara, N., Gerber, H. P. & LeCouter, J. The biology of VEGF and its receptors. *Nature Med.* 9, 669–676 (2003).
2. Ferrara, N., Mass, R. D., Campa, C. & Kim, R. Targeting VEGF-A to treat cancer and age-related macular degeneration. *Annu. Rev. Med.* 58, 491–504 (2007).
3. Ingber, D. E. & Folkman, J. Mechanochemical switching between growth and differentiation during fibroblast growth factor-stimulated angiogenesis *in vitro*: role of extracellular matrix. *J. Cell Biol.* 109, 317–330 (1989).
4. Chen, C. S., Mrksich, M., Huang, S., Whitesides, G. M. & Ingber, D. E. Geometric control of cell life and death. *Science* 276, 1425–1428 (1997).
5. Dike, L. E. *et al.* Geometric control of switching between growth, apoptosis, and differentiation during angiogenesis using micropatterned substrates. *In Vitro Cell. Dev. Biol. Anim.* 35, 441–448 (1999).
6. Parker, K. K. *et al.* Directional control of lamellipodia extension by constraining cell shape and orienting cell tractional forces. *FASEB J.* 16, 1195–1204 (2002).
7. Matthews, B. D., Overby, D. R., Mannix, R. & Ingber, D. E. Cellular adaptation to mechanical stress: role of integrins, Rho, cytoskeletal tension and mechanosensitive ion channels. *J. Cell Sci.* 119, 508–518 (2006).
8. Kumar, S. *et al.* Viscoelastic retraction of single living stress fibers and its impact on cell shape, cytoskeletal organization, and extracellular matrix mechanics. *Biophys. J.* 90, 3762–3773 (2006).
9. Moore, K. A. *et al.* Control of basement membrane remodeling and epithelial branching morphogenesis in embryonic lung by Rho and cytoskeletal tension. *Dev. Dyn.* 232, 268–281 (2005).
10. Huang, S. & Ingber, D. E. The structural and mechanical complexity of cell-growth control. *Nature Cell Biol.* 1, E131–E138 (1999).
11. Folkman, J. & Moscona, A. Role of cell shape in growth control. *Nature* 273, 345–349 (1978).
12. Folkman, J. & Kalluri, R. Cancer without disease. *Nature* 427, 787 (2004).
13. Matsumoto, T. & Claesson-Welsh, L. VEGF receptor signal transduction. *Sci. STKE* 2001, re21 (2001).
14. Wong, C. G., Rich, K. A., Liaw, L. H., Hsu, H. T. & Berns, M. W. Intravitreal VEGF and bFGF produce florid retinal neovascularization and hemorrhage in the rabbit. *Curr. Eye Res.* 22, 140–147 (2001).
15. Mammoto, A., Huang, S., Moore, K., Oh, P. & Ingber, D. E. Role of RhoA, mDia, and ROCK in cell shape-dependent control of the Skp2-p27kip1 pathway and the G1/S transition. *J. Biol. Chem.* 279, 26323–26330 (2004).
16. Mammoto, A., Huang, S. & Ingber, D. E. Filamin links cell shape and cytoskeletal structure to Rho regulation by controlling accumulation of p190RhoGAP in lipid rafts. *J. Cell Sci.* 120, 456–467 (2007).
17. Jiang, W. *et al.* An FF domain-dependent protein interaction mediates a signaling pathway for growth factor-induced gene expression. *Mol. Cell* 17, 23–35 (2005).
18. Jackson, T. A., Taylor, H. E., Sharma, D., Desiderio, S. & Danoff, S. K. Vascular endothelial growth factor receptor-2: counter-regulation by the transcription factors, TFII-I and TFII-IRD1. *J. Biol. Chem.* 280, 29856–29863 (2005).
19. Roy, A. L. Biochemistry and biology of the inducible multifunctional transcription factor TFII-I. *Gene* 274, 1–13 (2001).
20. Francke, U. Williams–Beuren syndrome: genes and mechanisms. *Hum. Mol. Genet.* 8, 1947–1954 (1999).
21. Roy, A. L. Signal-induced functions of the transcription factor TFII-I. *Biochim. Biophys. Acta* 1769, 613–621 (2007).
22. Patterson, C. *et al.* Cloning and functional analysis of the promoter for KDR/flk-1, a receptor for vascular endothelial growth factor. *J. Biol. Chem.* 270, 23111–23118 (1995).
23. Minami, T., Rosenberg, R. D. & Aird, W. C. Transforming growth factor-β_1-mediated inhibition of the *flk-1/KDR* gene is mediated by a 5′-untranslated region palindromic GATA site. *J. Biol. Chem.* 276, 5395–5402 (2001).
24. Minami, T. *et al.* Interaction between hex and GATA transcription factors in vascular endothelial cells inhibits *flk-1/KDR*-mediated vascular endothelial growth factor signaling. *J. Biol. Chem.* 279, 20626–20635 (2004).
25. Cantor, A. B. & Orkin, S. H. Hematopoietic development: a balancing act. *Curr. Opin. Genet. Dev.* 11, 513–519 (2001).
26. Grogan, J. L. & Locksley, R. M. T helper cell differentiation: on again, off again. *Curr. Opin. Immunol.* 14, 366–372 (2002).
27. Pai, S. Y., Truitt, M. L. & Ho, I. C. GATA-3 deficiency abrogates the development and maintenance of T helper type 2 cells. *Proc. Natl Acad. Sci. USA* 101, 1993–1998 (2004).
28. Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D. & Werb, Z. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* 127, 1041–1055 (2006).
29. Su, Z. J. *et al.* A vascular cell-restricted RhoGAP, p73RhoGAP, is a key regulator of angiogenesis. *Proc. Natl Acad. Sci. USA* 101, 12212–12217 (2004).
30. Arthur, W. T., Petch, L. A. & Burridge, K. Integrin engagement suppresses RhoA activity via a c-Src-dependent mechanism. *Curr. Biol.* 10, 719–722 (2000).
31. Robinson, C. J. & Stringer, S. E. The splice variants of vascular endothelial growth factor (VEGF) and their receptors. *J. Cell Sci.* 114, 853–865 (2001).
32. Sheibani, N. & Frazier, W. A. Down-regulation of platelet endothelial cell adhesion molecule-1 results in thrombospondin-1 expression and concerted regulation of endothelial cell phenotype. *Mol. Biol. Cell* 9, 701–713 (1998).
33. Numaguchi, Y. *et al.* Caldesmon-dependent switching between capillary endothelial cell growth and apoptosis through modulation of cell shape and contractility. *Angiogenesis* 6, 55–64 (2003).
34. Polte, T. R., Eichler, G. S., Wang, N. & Ingber, D. E. Extracellular matrix controls myosin light chain phosphorylation and cell contractility through modulation of cell shape and cytoskeletal prestress. *Am. J. Physiol. Cell Physiol.* 286, C518–C528 (2004).

35. Pierce, E. A., Avery, R. L., Foley, E. D., Aiello, L. P. & Smith, L. E. Vascular endothelial growth factor/vascular permeability factor expression in a mouse model of retinal neovascularization. *Proc. Natl Acad. Sci. USA* **92**, 905–909 (1995).

36. Stalmans, I. *et al.* Arteriolar and venular patterning in retinas of mice selectively expressing VEGF isoforms. *J. Clin. Invest.* **109**, 327–336 (2002).

37. Mammoto, T. *et al.* Angiopoietin-1 requires p190RhoGAP to protect against vascular leakage *in vivo*. *J. Biol. Chem.* 282, 23910–23918 (2007).

38. Singh, H., Medina, K. L. & Pongubala, J. M. Contingent gene regulatory networks and B cell fate specification. *Proc. Natl Acad. Sci. USA* **102**, 4949–4953 (2005).

39. Swiers, G., Patient, R. & Loose, M. Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. *Dev. Biol.* **294**, 525–540 (2006).

40. Gottgens, B. *et al.* Establishing the transcriptional programme for blood: the SCL stem cell enhancer is regulated by a multiprotein complex containing Ets and GATA factors. *EMBO J.* **21**, 3039–3050 (2002).

41. Engler, A. J., Sen, S., Sweeney, H. L. & Discher, D. E. Matrix elasticity directs stem cell lineage specification. *Cell* **126**, 677–689 (2006).

42. Clark, E. R. & Clark, E. L. Microscopic observations on the growth of blood capillaries in the living mammal. *Am. J. Anat.* **64**, 251–301 (1939).

43. Carmeliet, P. & Tessier-Lavigne, M. Common mechanisms of nerve and blood vessel wiring. *Nature* **436**, 193–200 (2005).

METHODS

Materials. Anti-GATA2 polyclonal antibody was from Abcam, anti-HA monoclonal antibody was from Covance, monoclonal antibodies against TFII-I, p190RhoGAP, PECAM (CD31) and paxillin were from Transduction laboratory, anti-GAPDH antibody was from Chemicon, anti-lamin monoclonal antibody was from Upstate, anti-VEGFR2 antibody was from Cell Signaling, and anti-Myc antibody was from Santa Cruz. Protein G-sepharose was from Amersham-Pharmacia and SU5416 was from Calbiochem. VEGF-A was from NIH; bFGF and PDGF were from Roche and Biovision, respectively. Cell permeable Rho inhibitor (C3 exoenzyme) was from Cytoskeleton. HMVE and HUVE cells (Cambrex) were cultured in EBM2 medium containing 5% FBS and growth factors (VEGF, bFGF and PDGF)[16] for all experiments except the nuclear translocation assays in which we used EBM2 with 0.3% serum. Cells were plated on plastic dishes for molecular biochemical assays, and on fibronectin-coated glass coverslips for cell staining, except for experiments using flexible substrates.

Plasmid construction and gene knockdown. pGL3-VEGFR2-225 (-225 to $+268$), -570 (-570 to $+268$), -780 (-780 to $+268$) were constructed using RT–PCR with genomic DNA from HUVE cells, and subcloned into pGL3 vector (Promega) at the SacI/XhoI sites. For pGL3-VEGFR2-780Inr-MUT, the Inr (CACT to GTGC) was point mutated using the QuickChange mutagenesis kit (Stratagene). For lentivirus construction, human *Myc-TFII-I* (delta isoform), *HA-GATA2* and *HA-p190RhoGAP* were constructed by PCR using template plasmids from Open Biosystems and H. Sabe (Osaka Bioscience Institute). For retrovirus construction, mouse *Myc-TfII-I* (gamma isoform) and *HA-Gata2* were constructed using template plasmids from Open Biosystems and T. Nakano (Osaka University), respectively. To generate the delta isoform of mouse TFII-I, 256–274 and 294–314 amino acids were deleted from the gamma isoform. Construction of constitutively active RhoA and the generation of viral vectors were previously described[37]. For gene knockdown, siRNA transfection was performed as described[16]. siRNA sequences are shown in Supplementary Table 1.

Biochemical methods. For luciferase reporter assays, HUVE cells were transfected using Superfect (QIAGEN) and assayed using Dual-Luciferase reporter assay kit (Promega). Luciferase activity was measured in duplicate using a luminometer (TD20/20, Turner Designs). Cytoplasmic and nuclear cell extracts were prepared with a Nuclear Extraction Kit (Chemicon).

Molecular biological methods. Quantitative RT–PCR was performed with the Quantitect SYBR Green RT–PCR kit (QIAGEN) using ABI7300 real-time PCR system (Applied Biosystems); β2 microglobulin or cyclophilin controlled for complementary DNA content. The primers used are shown in Supplementary Table 2. For ChIP assay, DNA from HMVE cells transfected with control or *TFII-I* siRNA was immunoprecipitated with the GATA2 antibody or control immunoglobulin (Jackson Immuno Research), or vice versa, according to the manufacturer instructions (Active Motif). The GATA2- and TFII-I-binding region was amplified using primers, 5′-GTAAATGGGCTTGGGGAGCTG-3′ and 5′-GGCG-GCTGCAGGGGCGTCT-3′.

Cell analysis methods. Flexible polyacrylamide gel culture substrates were prepared as described[34,44] and coated with fibronectin (1 µg cm^{-2}). Substrate flexibility was controlled by varying the acrylamide (2–4%) and the bis-acrylamide (0.1–0.5%) concentration; the Young's modulus (stiffness) was determined as described[45]. HMVE cells were cultured for 6 h on the gels and immunostaining was performed and analysed using confocal Leica SP2 microscope[15]. For cell migration assay, transwell membranes (Coster) were coated with 0.5% gelatin, and cells were seeded (10^5 cells per 100 µl) with 0.3% FBS in EBM2. Cells were stained with Giemsa solution 16 h later, and counted in ten random fields ($\times400$). For the *in vitro* angiogenesis assay, HMVE cells (10^4 cells per 150 µl of EBM-2) were plated on Matrigel (BD Biosciences), and incubated for 12–16 h in the presence of VEGF (10 ng ml^{-1}); tube formation was assessed in ten random fields ($\times4$).

***In vivo* Matrigel implantation assay.** All animal studies were reviewed and approved by the Animal Care and Use Committee of Children's Hospital Boston. Matrigel plugs with different elasticity were cast in 4×4 mm (internal diameter $\times$ height) polydimethylsiloxane moulds and incubated at 37 °C overnight before implanting them subcutaneously on the backs of C57BL/6 mice. The stiffness of the Matrigel was modulated over a narrow range (that is, without making it rigid) by altering ECM protein cross-linking using a microbial transglutaminase (2.5–20 U g^{-1}; Ajinomoto)[46]. The storage modulus (G') of the gels was measured with an AR-G2 rheometer (TA Instruments) using a standard 20-mm aluminium parallel plate (1 Hz, 1% strain, 37 °C). The Young's modulus for an equivalent polyacrylamide gel was calculated by $E = 2G'(1 + v)$ using an average Poisson's ratio (v) of 0.5. After 7 days, the polydimethylsiloxane moulds containing the gel plugs were collected, fixed and cryosectioned. H&E staining and immunostaining were performed as described[15,47]. Stacks of optical sections (20-µm thick) were compiled to form three-dimensional images using Volocity4.4 (Improvision, PerkinElmer). Vessel formation was evaluated by counting the number of vessels that stained positive for fluorescein-conjugated ConA injected into the tail vein or VEGFR2 in five different areas ($n = 6$), individual cell nuclei were identified by DAPI staining. Cell recovery solution (BD Biosciences) was used to collect cells from the recovered Matrigel plugs. In some studies, siRNA (7 µg) was mixed into the Matrigels and 10 µg of additional siRNA was injected into the implanted Matrigel after 3 days ($n = 6$); the same amount of scrambled siRNA was used as a control. Gene knockdown was evaluated by counting the number of cells expressing each gene in the five different regions.

***In vivo* analysis for retinal vessel formation.** For gene knockdown in living retina, siRNA (0.5 µg) for each gene was injected intravitreally into one eye of a C57BL/6 mouse at P5 or P14, and the same amount of control siRNA was injected into the other eye. To overexpress genes, the complex of DNA (0.5 µg) for *TFII-I* and/or *GATA2* and jetPEI transfection reagent (Polyplus transfection) was injected into the eye at P14. Vascular network formation in the retina was assessed 2 days after injection using flat-mounted, fluorescein-conjugated isolectin-staining and immunohistochemical analysis ($n = 7$). Retinal RNA was purified and gene expression was quantified using qRT–PCR ($n = 7$). Quantification of vessel density was performed with Adobe Photoshop.

44. Pelham, R. J. Jr & Wang, Y. Cell locomotion and focal adhesions are regulated by substrate flexibility. *Proc. Natl Acad. Sci. USA* **94**, 13661–13665 (1997).
45. Wang, N. *et al.* Cell prestress. I. Stiffness and prestress are closely associated in adherent contractile cells. *Am. J. Physiol. Cell Physiol.* **282**, C606–C616 (2002).
46. Yung, C. W. *et al.* Transglutaminase crosslinked gelatin as a tissue engineering scaffold. *J. Biomed. Mater. Res. A* **83**, 1039–1046 (2007).
47. Connor, K. M. *et al.* Increased dietary intake of Ω-3-polyunsaturated fatty acids reduces pathological retinal angiogenesis. *Nature Med.* **13**, 868–873 (2007).

LETTERS

A record of planet migration in the main asteroid belt

David A. Minton[1] & Renu Malhotra[1]

The main asteroid belt lies between the orbits of Mars and Jupiter, but the region is not uniformly filled with asteroids. There are gaps, known as the Kirkwood gaps, in distinct locations that are associated with orbital resonances with the giant planets[1]; asteroids placed in these locations will follow chaotic orbits and be removed[2]. Here we show that the observed distribution of main belt asteroids does not fill uniformly even those regions that are dynamically stable over the age of the Solar System. We find a pattern of excess depletion of asteroids, particularly just outward of the Kirkwood gaps associated with the 5:2, the 7:3 and the 2:1 Jovian resonances. These features are not accounted for by planetary perturbations in the current structure of the Solar System, but are consistent with dynamical ejection of asteroids by the sweeping of gravitational resonances during the migration of Jupiter and Saturn ~4 Gyr ago.

The Kirkwood gaps have been explained by the perturbing effects of the giant planets that cause dynamical chaos and orbital instabilities on very long timescales in narrow zones in the main asteroid belt[2], but thus far it has not been established how much of the main belt asteroid distribution is accounted for by planetary perturbations alone. We compared the distribution of observed asteroids against a model asteroid belt uniformly populated in the dynamically stable zones. Our model asteroid belt was constructed as follows. Test particle asteroids were given eccentricity and inclination distributions similar to the observed main belt, but a uniform distribution in semimajor axis. We then performed a numerical integration for 4 Gyr of the test particles' orbital evolution under the gravitational perturbations of the planets using a parallelized implementation of a symplectic mapping[3,4]. Details of the simulation can be found in Supplementary Information.

We sorted the surviving particles into semimajor axis bins of width 0.03 AU. We compared the model asteroid belt with the observed asteroid belt, as shown in Fig. 1a. We find that the observed asteroid belt is overall more depleted than the model can account for, and there is a particular pattern in the excess depletion (Fig. 1a): there is enhanced depletion in the semimajor ranges spanning 2.81–3.11 AU and 3.34–3.47 AU; these regions are just exterior to the major Kirkwood gaps associated with the 5:2, 7:3 and 2:1 mean motion resonances (MMRs) with Jupiter; the semimajor axis ranges 2.72–2.81 AU and 3.11–3.23 AU, which are just interior to the 5:2 and the 2:1 resonances, do not show depletion. In addition the inner belt region, spanning 2.21–2.72 AU, shows significant excess depletion.

The above conclusions about the patterns of depletion are based on our model asteroid belt, which assumes uniform initial population of the dynamically stable zones. It is conceivable that the discrepancies between the model and the observations could be due to a non-uniform initial distribution of asteroids. However, the particular features we find cannot be explained by appealing to the primordial distribution of planetesimals in the solar nebula, nor to the effects of the mass depletion that occurred during the planet formation era (see Supplementary Information). As we show below, they can instead be readily accounted for by the effects of giant planet migration in the early history of the Solar System.
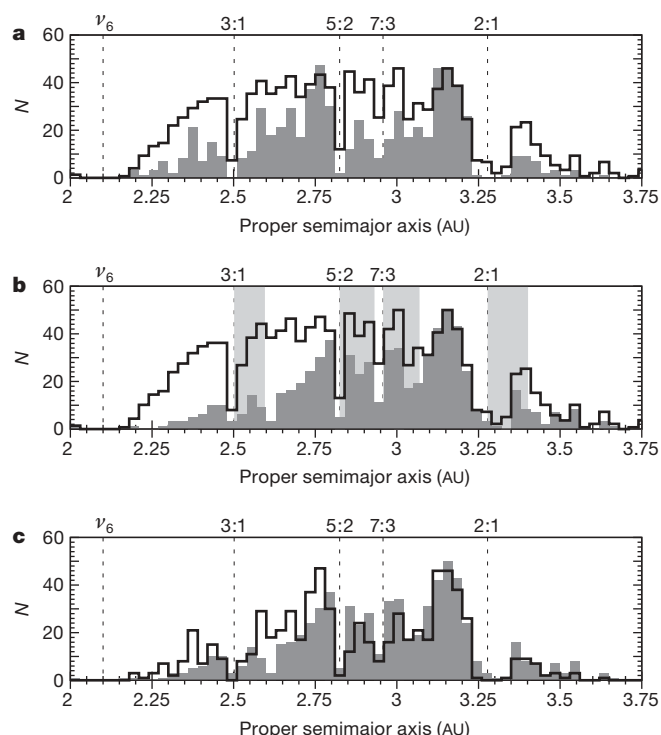
Figure 1 | Comparison of the observed main belt asteroid distribution with our simulated asteroid belt and results of the migration simulation. a, Solid line histogram, the distribution of asteroids remaining in our model asteroid belt at the end of the 4 Gyr simulation in which the asteroid belt region was initially uniformly populated with test particles and the planets were in their current orbits. Shaded histogram, our observational comparison sample, which consists of the 690 asteroids with absolute magnitude $H < 9.7$ (equivalent to diameters $D \gtrsim 50$ km, assuming a visual geometric albedo of 0.09), in the AstDys online data service[19] (details in Supplementary Information). The model asteroid belt (solid line) was normalized by multiplying all bins by a constant such that the value of the most-populous model bin equalled that of its corresponding bin in the observations. Vertical dashed lines, current positions of the ν_6 secular resonance and the strong Jovian MMRs associated with the major Kirkwood gaps. **b**, Solid line, initial distribution of test particles in the simulation with migrating planets. Shaded histogram, the distribution of test particles remaining at the end of the 100 Myr migration simulation. The post-migration test particle bins were normalized by multiplying all bins by a constant value such that the value of the most-populous bin equalled that of its corresponding bin in the initial conditions. The planet migration history followed the form of equation (1). Grey shading, regions swept by the strong Jovian MMRs. **c**, Comparison of the model asteroid belt subjected to planet migration and the observed asteroid belt. Solid line, distribution of observed large asteroids (the same as the shaded histogram in **a**). Shaded histogram, distribution of test particles remaining at the end of the 100 Myr migration simulation (the same as the shaded histogram in **b**).

[1]Lunar and Planetary Laboratory, University of Arizona, 1629 East University Boulevard, Tucson, Arizona 85716, USA.

There is evidence in the outer Solar System that the giant planets—Jupiter, Saturn, Uranus and Neptune—did not form where we find them today. The orbit of Pluto and other Kuiper belt objects that are trapped in MMRs with Neptune can be explained by the outward migration of Neptune due to interactions with a more massive primordial planetesimal disk in the outer regions of the Solar System[5,6]. The exchange of angular momentum between planetesimals and the four giant planets caused the orbital migration of the giant planets until the outer planetesimal disk was depleted of most of its mass, leaving the giant planets in their present orbits[7–9]. As Jupiter and Saturn migrated, the locations of MMRs and secular resonances swept across the asteroid belt, exciting asteroids into terrestrial-planet-crossing orbits, thereby greatly depleting the asteroid belt population and perhaps also causing a late heavy bombardment in the inner Solar System[10–13].

We performed a computer simulation to test the hypothesis that the patterns of asteroid depletion inferred from Fig. 1a are consistent with planet migration. We used a total of 1,819 surviving test particles from the previous 4 Gyr simulation as initial conditions for a simulation with migrating planets. For the purposes of this simulation, we applied an external tangential force on each of the planets to simulate their orbital migration, so that a planet's semimajor axis evolved as follows[5]:

$$a(t) = a_0 + \Delta a[1 - \exp(-t/\tau)] \qquad (1)$$

where a_0 is the initial semimajor axis, Δa is the migration distance, and τ is a migration rate e-folding time. Jupiter, Saturn, Uranus and Neptune had initial semimajor axes displaced from their current values by $\Delta a = +0.2$, -0.8, -3.0 and -7.0 AU, respectively; these values are consistent with other estimates of Jupiter's and Neptune's migration distances[6,7,9,14], but Uranus' and Saturn's migration distances are less certain. We used $\tau = 0.5$ Myr, which is near the lower limit inferred from Kuiper belt studies[15]. After 100 Myr of evolution under the influence of migrating planets, the 687 surviving test particles in the simulation were sorted and binned. The distribution of the survivors is shown in Fig. 1b.

We see directly that, in contrast with Fig. 1a, the asteroid belt distribution produced by the planet migration model matches qualitatively quite well the distribution of the observed asteroids (Fig. 1c). The patterns of excess depletion that we noted in Fig. 1a are explained well by the effects of the orbital migration of Jupiter and Saturn: the regions within the sweep zones of the 5:2, 7:3 and 2:1 Jovian MMRs show enhanced depletion, and the migration model also accounts for the excess depletion in the inner belt.

Of note is that the inner asteroid belt region (2.15–2.81 AU) is somewhat more depleted in the migration simulation than in the observations. The majority of depletion from this region found in our migration simulation is due to the sweeping v_6 secular resonance. This powerful resonance removes asteroids from the main belt by secularly increasing their eccentricities to planet-crossing values[16]. The maximum eccentricity of an asteroid disturbed by the passage of the v_6 resonance, and thereby the degree of asteroid depletion, is related to the sweeping speed: the slower the sweeping, the more the depletion[17]. The distances the planets migrate determine the ranges in asteroids' semimajor axes that are affected by the sweeping. In our simulation, as Jupiter and Saturn migrated, the v_6 secular resonance swept inward across the entire main asteroid belt to its present location at ~ 2.1 AU, as shown in Fig. 2. But because the v_6 resonance location is such a steep function of Saturn's semimajor axis (see Fig. 2), even modest proposed values of Saturn's migration distance result in all of the asteroid belt being affected by the passage of this resonance. Thus, the overall level of depletion of the asteroid belt is most strongly dependant on the speed of planet migration, and only secondarily on the migration distance. Because we used an exponentially decaying migration rate for the giant planets, the v_6 resonance sweeping rate decreased as it approached its current location, thereby causing relatively greater asteroid depletion in the inner belt. Thus,
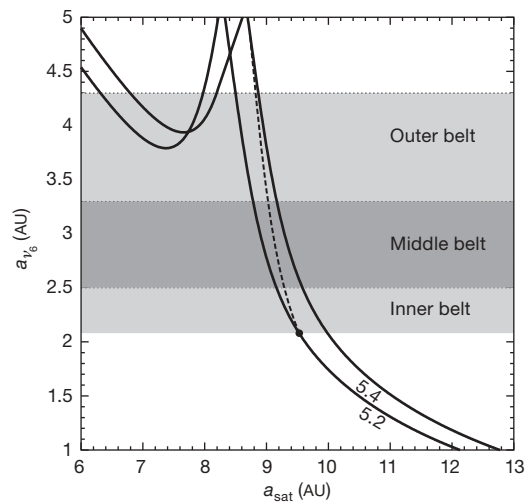
Figure 2 | The location of the v_6 secular resonance as a function of Saturn's semimajor axis. This is calculated using linear secular theory, with a correction for the effects of the $(n + 1):n$ MMRs, up to and including $n = 8$, between Jupiter and Saturn[20]. A dot shows the current location of the v_6. Labelled solid lines, calculations based on different fixed values of Jupiter's semimajor axis (given in AU). Dashed line, the path of the v_6 resonance in our migration simulation in which Jupiter and Saturn migrated -0.2 and $+0.8$ AU, respectively.

the small but noticeable differences between the model and the observations in Fig. 1c are sensitive to the details of the time history of the planet migration speed.

We note that our model asteroid belt lost 62% of its initial premigration population, but the actual asteroid belt may have lost as much as ~90–95% of its asteroids by migration[18]. Because the overall level of asteroid depletion is particularly sensitive to the speed of planet migration, detailed exploration of the parameters of the planet migration model and comparison with observations of main belt asteroids may provide strong quantitative constraints on planet migration.

1. Kirkwood, D. *Meteoric Astronomy: A Treatise on Shooting-stars, Fireballs, and Aerolites* (Lippincott, 1867).
2. Wisdom, J. Chaotic behaviour in the solar system. *Proc. R. Soc. Lond. A* **413**, 109–129 (1987).
3. Wisdom, J. & Holman, M. Symplectic maps for the n-body problem. *Astron. J.* **102**, 1528–1538 (1991).
4. Saha, P. & Tremaine, S. Symplectic integrators for solar system dynamics. *Astron. J.* **104**, 1633–1640 (1992).
5. Malhotra, R. The origin of Pluto's peculiar orbit. *Nature* **365**, 819–821 (1993).
6. Malhotra, R. The origin of Pluto's orbit: Implications for the solar system beyond Neptune. *Astron. J.* **110**, 420–429 (1995).
7. Fernandez, J. A. & Ip, W.-H. Some dynamical aspects of the accretion of Uranus and Neptune – The exchange of orbital angular momentum with planetesimals. *Icarus* **58**, 109–120 (1984).
8. Hahn, J. M. & Malhotra, R. Orbital evolution of planets embedded in a planetesimal disk. *Astron. J.* **117**, 3041–3053 (1999).
9. Tsiganis, K., Gomes, R., Morbidelli, A. & Levison, H. F. Origin of the orbital architecture of the giant planets of the Solar System. *Nature* **435**, 459–461 (2005).
10. Liou, J.-C. & Malhotra, R. Depletion of the outer asteroid belt. *Science* **275**, 375–377 (1997).
11. Levison, H. F. et al. Could the lunar "Late Heavy Bombardment" have been triggered by the formation of Uranus and Neptune? *Icarus* **151**, 286–306 (2001).
12. Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* **435**, 466–469 (2005).
13. Strom, R. G., Malhotra, R., Ito, T., Yoshida, F. & Kring, D. A. The origin of planetary impactors in the inner solar system. *Science* **309**, 1847–1850 (2005).
14. Franklin, F. A., Lewis, N. K., Soper, P. R. & Holman, M. J. Hilda asteroids as possible probes of Jovian migration. *Astron. J.* **128**, 1391–1406 (2004).
15. Murray-Clay, R. A. & Chiang, E. I. A signature of planetary migration: The origin of asymmetric capture in the 2:1 resonance. *Astrophys. J.* **619**, 623–638 (2005).
16. Murray, C. D. & Dermott, S. F. *Solar System Dynamics* (Cambridge Univ. Press, 1999).

17. Heppenheimer, T. A. Secular resonances and the origin of eccentricities of Mars and the asteroids. *Icarus* **41**, 76–88 (1980).
18. O'Brien, D. P., Morbidelli, A. & Bottke, W. F. The primordial excitation and clearing of the asteroid belt—Revisited. *Icarus* **191**, 434–452 (2007).
19. Knežević, Z. & Milani, A. Proper element catalogs and asteroid families. *Astron. Astrophys.* **403**, 1165–1173 (2003).
20. Malhotra, R., Fox, K., Murray, C. D. & Nicholson, P. D. Secular perturbations of the Uranian satellites – Theory and practice. *Astron. Astrophys.* **221**, 348–358 (1989).

LETTERS

Large positive magnetoresistive effect in silicon induced by the space-charge effect

Michael P. Delmo[1], Shinpei Yamamoto[2], Shinya Kasai[1], Teruo Ono[1] & Kensuke Kobayashi[1]

Recent discoveries of large magnetoresistance in non-magnetic semiconductors[1–8] have gained much attention because the size of the effect is comparable to, or even larger than, that of magnetoresistance in magnetic systems[9–14]. Conventional magnetoresistance in doped semiconductors is straightforwardly explained as the effect of the Lorentz force on the carrier motion[15], but the reported unusually large effects imply that the underlying mechanisms have not yet been fully explored. Here we report that a simple device, based on a lightly doped silicon substrate between two metallic contacts, shows a large positive magnetoresistance of more than 1,000 per cent at room temperature (300 K) and 10,000 per cent at 25 K, for magnetic fields between 0 and 3 T. A high electric field is applied to the device, so that conduction is space-charge limited[16–18]. For substrates with a charge carrier density below $\sim 10^{13}$ cm^{-3}, the magnetoresistance exhibits a linear dependence on the magnetic field between 3 and 9 T. We propose that the observed large magnetoresistance can be explained by quasi-neutrality breaking of the space-charge effect, where insufficient charge is present to compensate the electrons injected into the device. This introduces an electric field inhomogeneity, analogous to the situation in other semiconductors in which a large, non-saturating magnetoresistance was observed[1–5,19]. In this regime, the motions of electrons become correlated, and thus become dependent on magnetic field. Although large positive magnetoresistance at room temperature has been achieved in metal–semiconductor hybrid devices[6–8], we have now realized it in a simpler structure and in a way different from other known magnetoresistive effects[9–14,20]. It could be used to develop new magnetic devices from silicon, which may further advance silicon technology.

We used a lightly phosphorus-doped, n-type-silicon (n-Si) substrate. Figure 1a shows the temperature dependence of the mobility, μ, and the carrier density, n, of the substrate, whose behaviours are just as the standard theory predicts. The In/n-Si/In devices (In, indium) were fabricated by making indium ohmic contacts on the substrate (Fig. 1b; L is the distance between the contacts and W is their width). We measured the device current, I, between the contacts at the bias voltage, V, in a current-in-plane geometry with the magnetic field, H, applied perpendicular to the substrate.

Figure 1c shows typical I–V characteristics of one of the devices (sample 24) for various magnetic fields at 25 K, where $n = 2.7 \times 10^9$ cm^{-3} and $\mu = 2.0$ m^2 V^{-1} s^{-1}. The I–V curve for $H = 0$ T is almost linear ($I \propto V$) at $V < 1$ V, whereas there is a crossover to a nonlinear regime at larger values of V, which is a clear signature that the device is in the space-charge regime. As V increases, the injected carrier density, n_{inj}, starts to exceed the original carrier density, n, causing electrons to accumulate around the injection contact[16–18]. Instead of Ohm's law, the electric field induced by the space charge determines the current so as to satisfy continuity of current

and Poisson's equations. The high mobility and low carrier density favour this phenomenon. Actually, the I–V curve at $|V| > 3$ V obeys the Mott–Gurney law, given by $J = 9\varepsilon\mu V^2/8L^3$, which holds for unipolar space-charge-limited transport[16–18] (J is the current density and ε is the dielectric constant of silicon). The simulated I–V curve based on the Mott–Gurney law calculated using the geometric parameters is shown by the dotted black curve in Fig. 1c. The injected carrier
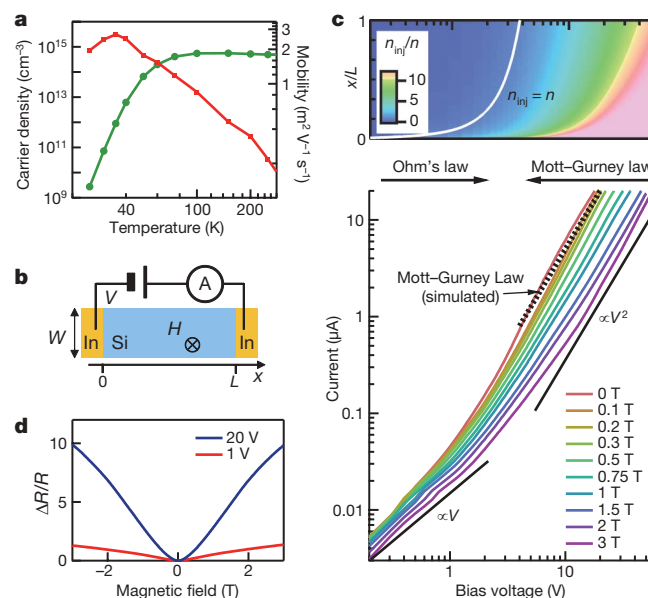
Figure 1 | Magnetoresistance of the In/n-Si/In device at 25 K. a, Carrier density (green curve; left-hand scale) and mobility (red curve; right-hand scale) as functions of temperature in the n-Si substrate. **b,** Sketch of the device. L is the distance between the contacts and W is their width. The magnetic field, H, is applied perpendicular to the current direction. **c,** I–V characteristics of the fabricated In/n-Si/In device (sample 24) for a variety of applied magnetic fields from $H = 0$ to 3 T at 25 K, plotted on a double-logarithmic scale. This device is fabricated on the 5 mm × 5 mm substrate of thickness 525 μm, and the indium contacts are about 0.3 mm × 0.3 mm with $L = 0.8$ mm. The two black lines below the I–V curves in the bottom panel show the slope of the ideal I–V curves for ohmic ($I \propto V$) and Mott–Gurney ($I \propto V^2$) behaviour (the lines are guides to the eye). The dotted line is the simulated I–V curve in the Mott–Gurney regime calculated using the geometrical parameters. The top panel shows n_{inj} calculated as a function of both position, x, normalized to L, and the bias voltage. At low bias voltage, where n_{inj} is much less than the equilibrium electron density, n, the distribution of the total charge carriers ($n + n_{inj}$) in the material is uniform. As the bias voltage increases, the charge distribution becomes non-uniform and is dominated by n_{inj}. The white line shows the boundary $n_{inj} = n$. **d,** Magnetoresistance normalized by the zero-field resistance, R, at given bias voltage, shown for the ohmic regime (red line) and for the Mott–Gurney regime (blue) at $T = 25$ K.

[1]Institute for Chemical Research, [2]Institute for Integrated Cell-Material Sciences, Kyoto University, Uji 611-0011, Japan.

density calculated in the Mott–Gurney model is shown in the upper panel of Fig. 1c as a function of both distance, x, from the injection contact and bias voltage.

As the perpendicular magnetic field increases from 0 T, the current at a fixed bias voltage monotonically decreases as shown in Fig. 1c, resulting in the positive magnetoresistance ratio

$$\mathrm{MR}(H) = \frac{R(H) - R(H=0)}{R(H=0)} = \frac{\Delta R}{R}$$

Here $R(H)$ is the two-terminal resistance (V/I) at magnetic field H. Surprisingly, MR changes between the ohmic regime and the Mott–Gurney regime; the I–V characteristics are more sensitive to the magnetic field in the latter regime than in the former regime. As shown in Fig. 1d, in the Mott–Gurney regime ($V = 20$ V) MR is ~1,000% at 3 T, which is about ten times larger than that in the ohmic regime ($V = 1$ V). The magnetoresistance is much reduced when the magnetic field is applied parallel to the current, indicating that the present effect results from the orbital magnetoresistance (Supplementary Fig. 1). The magnetoresistive effect in the Mott–Gurney regime was investigated in ref. 21, where the regime was realized only at high magnetic field above 3 T, unlike in the present case.

To confirm that the observed effect results from the electron transport through the silicon, the internal voltage differences (V_{OA}, V_{OB} and V_{OC}) between the voltage probes (O, A, B and C) were measured with the constant current applied to the device (sample 43) as shown in the top panel of Fig. 2a. When the current is as low as 0.1 μA, the electron transport obeys Ohm's law (dashed line in the bottom panel of Fig. 2a) because the voltage differences between the probes are linearly dependent on x, the distance of the probe from the injection contact. On the other hand, in the space-charge transport regime at $I = 40$ μA, the voltage differences is proportional to $x^{3/2}$ in agreement with the Mott–Gurney theory[16–18] that the electric field is spatially distributed in proportion to $x^{1/2}$ owing to the accumulation of the electrons around the injection contact.

The top panel of Fig. 2b represents the magnetoresistance ratio curve obtained for the two-terminal resistance of the same device. The observed MR is unusual in that after increasing quadratically up to ~3 T, it increases almost linearly between 3 and 9 T. This contradicts the naive expectation that MR in ordinary semiconductors saturates at sufficiently large magnetic field ($|H| \gg \mu^{-1} \approx 0.5$ T in this case)[15]. In the bottom panel we show the magnetoresistance ratios corresponding to the resistances R_{OA}, R_{OB} and R_{OC}, which are respectively defined by V_{OA}/I, V_{OB}/I and V_{OC}/I. Whereas R_{OA} saturates above 3 T, R_{OB} and R_{OC} increase almost linearly with the field up to 9 T, in agreement with the two-terminal magnetoresistance ratio. Because MR is larger between probes O and C than between probes O and B, the magnetoresistance must depend on the distance from the injection contact and, hence, on the local electric field. The present effect is thus unambiguously caused by the electron transport through the bulk silicon, and the electron orbital is proved to be more subject to the magnetic field in higher electric fields. We note that this unusual phenomenon is observed only below 40 K, where n decreases to less than ~10^{13} cm^{-3} (Supplementary Fig. 2).

A few mechanisms are known to induce large positive magnetoresistance in semiconductors. For example, lightly doped semiconductors show a similar effect in the hopping conduction regime at low temperatures owing to the shrinking of the wave function of the impurity state in the high magnetic field[22–24]. In our case, however, the temperature dependence of the resistivity and n ensures that the band transport is dominant at 25 K, where the mobility is as high as 2.0 m^2 V^{-1} s^{-1}. As a second example, positive magnetoresistance may occur by shorting of the Hall voltage as a result of the contact geometry when $L \ll W$ (Fig. 1b)[16,21]. Again, this does not explain the present effect because it is also observed in the long bar-shaped device ($L > W$).

The large positive magnetoresistance is also a characteristic of the inhomogeneous semiconductors[1–8,19,25–28]. In particular, the present distinctive magnetoresistance is reminiscent of the non-saturating, linear positive magnetoresistance caused by the current paths distorted in the magnetic field due to the inhomogeneity[1–5]. Although the theories[3,4,19,25–27] treat inherently disordered systems unlike the present lightly doped silicon, the observed position dependence of MR, which implies that the device is no longer homogeneous, together with the near linearity of MR, leads us to propose that the inhomogeneity induced in the Mott–Gurney regime by the broken quasi-neutrality is responsible for the present magnetoresistance. Indeed, the electron correlation is not negligible because, as $n_{inj} > n$, there are no compensating charges. Quantitatively, the Debye length[15,16] of our n-Si is 23 μm in equilibrium at 25 K. For $n_{inj}/n = 10$, the average distance between the injected electrons is 3.3 μm, indicating that the large numbers of electrons are correlated within the Debye length by means of the unscreened Coulomb interaction. The spatial distribution of n_{inj} and the electric field should be treated more directly beyond the mean-field treatment in the Mott–Gurney theory; when the electron trajectory is bent, the distortion in turn influences the trajectories of the surrounding electrons through the long-range interaction, which is expected to induce the local inhomogeneity in both the carrier density and the electric field. In such an unscreened regime, the randomness in the distribution of dopants may also enhance the induced inhomogeneity.

Given the inhomogeneity, it may yield[1,2,19] MR $\propto \mu H \approx 1,800\%$ at 9 T for $\mu = 2$ m^2 V^{-1} s^{-1}, which is consistent with the present observation. The magnetoresistance induced by the inhomogeneity can be characterized by the ratio between the average, $\langle \mu \rangle$, and the distribution width, $\Delta \mu$, of the local mobility[3,4]. The present magnetoresistance ratio
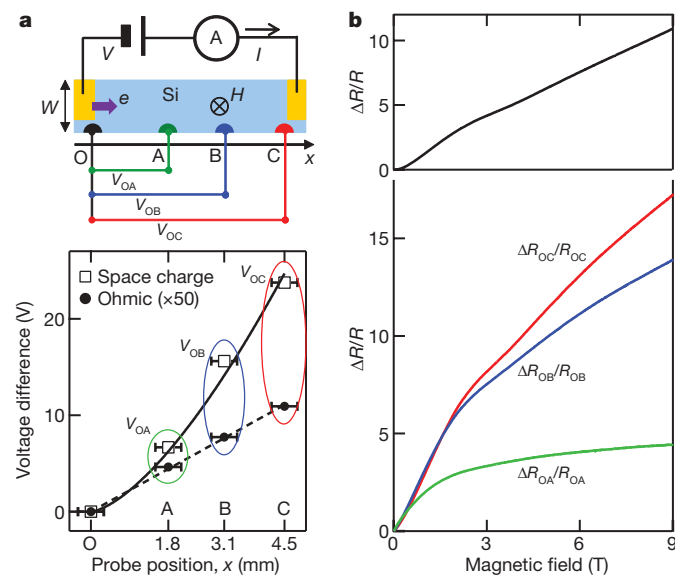

Figure 2 | Magnetoresistance probed by the internal voltage differences. a, Top: sketch of the In/n-Si/In device (sample 43), which has four voltage probes (O, A, B and C) on its lower side in addition to the injection contact on the left and the positive contact on the right. The distances between O and A, O and B, and O and C are respectively 1.8 mm, 3.1 mm and 4.5 mm. The width of the device is 2.8 mm. The width of each voltage probe is ~0.3 mm. Bottom: the measured voltage differences between O and A, O and B, and O and C, plotted for the ohmic case ($I = 0.1$ μA) and the Mott–Gurney case ($I = 40$ μA) for $H = 0$ T at 25 K. In the ohmic case, the voltage difference (multiplied by 50 for clarity) is a linear function of x, the distance of the probe from the injection contact, in agreement with the resistivity of the present silicon substrate. In the Mott–Gurney case, it is proportional to $x^{3/2}$, as expected. The error bars reflects the width of the voltage probe. **b**, Top: MR (magnetoresistance normalized by the zero-field resistance R) of sample 43 up to $H = 9$ T for $V = 50$ V at 25 K, obtained from the two-terminal resistance. Bottom: plots of MR for the zero-field resistances R_{OA}, R_{OB} and R_{OC}.

curve, with the crossover around 3 T, indicates that $\Delta\mu/\langle\mu\rangle$ is less than one.

On the basis of the above discussion, a similar effect is expected to occur at higher temperature only if the carrier density is small and the mobility is high. Figure 3a shows the I–V characteristics of the device (sample 32) made of the intrinsic silicon (i-Si) ($n = 1.78 \times 10^{12}$ cm^{-3} and $\mu = 1{,}300$ cm^2 V^{-1} s^{-1} at 300 K). Although the I–V characteristics at 300 K show ohmic behaviour at low bias voltages, the behaviour changes for $V > 50$ V, signalling the space-charge effect[16–18,29]. Again, a pronounced current suppression in the magnetic field occurs there. Figure 3b shows the large positive MR observed for $V > 50$ V and, for comparison, the usual small, quadratic MR observed at $V = 25$ and 50 V.

The mechanism causing the present effect is expected to work more effectively in higher electric fields. In Fig. 4a, the zero-field I–V curve shows the Mott–Gurney behaviour and then the breakdown at ~53 V as the electric field created by the bias voltage and the accumulated space charge exceed the threshold that the device material can sustain. As the magnetic field increases, breakdown is suppressed by the magnetic field making the device more resistive through the induced inhomogeneity, and the breakdown voltage is shifted to higher voltages. As a result, MR ≈ 10,000% for $V = 52$ V at 3 T just before breakdown. An identical phenomenon is observed in the i-Si device (sample 32) at 300 K, as shown in Fig. 4b, where MR exceeds ~1,000 at 3 T just before the breakdown at 71.2 V. In both cases, the magnetoresistance ratio curve around the breakdown voltage is not linear in H and is very sensitive to the bias voltage that determines the internal electric field.

The present space-charge-induced magnetoresistance should be observable in other materials with very low carrier densities and high mobilities. Two advantages of the effect in device applications are that the noise-equivalent field in the thermal noise limit is ~0.01 nT Hz$^{-1/2}$ (Fig. 4a) and that the space-charge transport favourably occurs at lower

a

b

Figure 4 | Large positive magnetoresistance at high electric field in n-Si and i-Si. a, At $H = 0$ T, the I–V curve of sample 24 shows a breakdown at 53 V above the Mott–Gurney regime at 25 K. As the magnetic field increases, the breakdown voltage shifts to higher voltage regimes. At $H = 3$ T, MR ≈ 10,000% at $V = 52$ V. The breakdown effect has been observed in several devices in the accessible voltage range, although the breakdown voltage is different among different devices. **b**, The I–V curve of the In/i-Si/In device at 300 K (sample 32) at $H = 0$, 0.5, 1 and 3 T. As observed in **a**, the breakdown voltage shifts to higher voltage when the magnetic field is applied. As a result, MR values exceeding ~1,000% are obtained between $H = 0$ and 3 T just before the breakdown at 71.2 V. The discontinuities in the I–V curves just before the breakdown may be caused by the partial breakdown at the imperfect ohmic contact. These discontinuities were often observed in the i-Si devices but not in the n-Si devices.

bias voltages for smaller devices. That the device resistance can be controlled by two parameters, namely the magnetic field and the bias voltage (Fig. 4b), is also of considerable technological importance. More quantitative understanding of this phenomenon will help in developing advanced silicon-based magnetoelectronics technologies. *Note added in proof:* Following acceptance of our paper, we became aware that Schoonus *et al.*[30] have achieved similarly large values of magnetoresistance in doped silicon at low temperatures. Their structures involved Schottky (rather than ohmic) contacts, in which impact ionization at the Schottky barrier is responsible for the observed magnetoresistance, as proposed in ref. 20. The mechanism we propose to explain our results (based on the space-charge effect) is different.

METHODS SUMMARY

Two different sorts of n-type-doped silicon substrates (thicknesses, 525 μm and 380 μm) and one intrinsic silicon substrate (thickness, 380 μm; Furuuchi Chemical Corporation) were used in this study. The nominal resistivities were 10–20 Ω cm for the n-type-doped silicon and >1,000 Ω cm for the intrinsic silicon. The 5 mm × 5 mm silicon samples were prepared for carrier density, n, and resistivity, ρ, measurements at various temperatures. The contact electrodes at the four corners of the square samples were prepared by soldering and alloying indium at 350 °C for five minutes. The carrier densities and resistivities of the n-Si substrates were measured using the standard van der Pauw method down to 25 K. The magnetic field was applied perpendicular to the substrate plane. The mobility, μ, was calculated from the experimentally obtained n and ρ ($\mu = 1/\rho q n$, q being carrier charge). The temperature dependences of n and μ for one of the n-Si substrates are shown in Fig. 1a. The Hall coefficient is constant between $H = -3$ and 3 T, which indicates that the equilibrium charge density of the silicon does not depend on H. We obtained the magnetoresistance result reported here in two different n-Si substrates.

a

b

Figure 3 | Large positive magnetoresistance of the intrinsic silicon at 300 K. a, I–V characteristics of the In/i-Si/In device (sample 35), shown for $H = 0$, 1 and 3 T. Clear deviation from ohmic transport is observed for $|V| > 50$ V. Note that the simple Mott–Gurney law is not applicable for i-Si because it is not unipolar at room temperature[29] (300 K). Inset, photo of the In/i-Si/In device (sample 35) with device dimensions. Its thickness is 380 μm. **b**, MR for various bias voltages. In the ohmic regime ($V = 25$ and 50 V), the device shows the usual small, quadratic magnetoresistance, whereas in the space-charge regime ($V = 100$, 150 and 200 V) the MR effect is pronounced.

In the two-terminal magnetotransport measurement, more than 40 devices with various contact geometries were made from n-Si and i-Si substrates (see Fig. 3a for an example). Good ohmic contacts were obtained by mechanically removing the naturally oxidized layer before making the indium contacts by alloying. The distance between the contacts ranged from 0.1 to 5 mm. In the internal voltage distribution measurement, a rectangular n-Si substrate with current injection and drain electrodes on one side and four voltage probe electrodes on the other side was fabricated (Fig. 2a). The ohmic contacts were made in the same way as those applied in the two-terminal devices. We measured voltage difference and magnetoresistance simultaneously at 25 K at various probe positions, as well as the two-terminal (injection–drain) magnetoresistance. The magnetic field was applied perpendicular to the substrate plane.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Xu, R. *et al.* Large magnetoresistance in non-magnetic silver chalcogenides. *Nature* **390,** 57–60 (1997).
2. Husmann, A. *et al.* Megagauss sensors. *Nature* **417,** 421–424 (2002).
3. Parish, M. M. & Littlewood, P. B. Non-saturating magnetoresistance in heavily disordered semiconductors. *Nature* **426,** 162–165 (2003).
4. Parish, M. M. & Littlewood, P. B. Classical magnetotransport of inhomogeneous conductors. *Phys. Rev. B* **72,** 094417 (2005).
5. Zhang, T. *et al.* Tuning the inherent magnetoresistance of InSb thin films. *Appl. Phys. Lett.* **88,** 012110 (2006).
6. Solin, S. A., Thio, T., Hines, D. R. & Heremans, J. J. Enhanced room-temperature geometric magnetoresistance in inhomogeneous narrow-gap semiconductor. *Science* **289,** 1530–1532 (2000).
7. Moussa, J. *et al.* Finite-element modelling of extraordinary magnetoresistance in thin film semiconductors with metallic inclusions. *Phys. Rev. B* **64,** 184410 (2001).
8. Solin, S. A. *et al.* Nonmagnetic semiconductors as read-head sensors for ultra-high-density magnetic recording. *Appl. Phys. Lett.* **80,** 4012–4014 (2002).
9. Baibich, M. N. *et al.* Giant magnetoresistance of Fe(001)/(001) Cr magnetic superlattices. *Phys. Rev. Lett.* **61,** 2472–2475 (1988).
10. Berkowitz, A. E. *et al.* Giant magnetoresistance in heterogeneous Cu-Co alloys. *Phys. Rev. Lett.* **68,** 3745–3748 (1992).
11. Xiao, J. Q., Jiang, J. & Chien, C. L. Giant magnetoresistance in nonmultilayer magnetic systems. *Phys. Rev. Lett.* **68,** 3749–3752 (1992).
12. Jin, S. *et al.* Thousandfold change in resistivity in magnetoresistive La-Ca-Mn-O films. *Science* **264,** 413–415 (1994).
13. Yuasa, S. *et al.* Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions. *Nature Mater.* **3,** 868–871 (2004).
14. Parkin, S. S. P. *et al.* Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers. *Nature Mater.* **3,** 862–867 (2004).
15. Ashcroft, N. W. & Mermin, N. D. *Solid State Physics* (Holt, Rinehart and Winston, 1976).
16. Sze, S. M. & Ng, K. K. *Physics of Semiconductor Devices* 3rd edn (Wiley, 2007).
17. Rose, A. Space-charge-limited currents in solids. *Phys. Rev.* **97,** 1538–1544 (1955).
18. Lampert, M. A. Simplified theory of space-charge-limited currents in an insulator with traps. *Phys. Rev.* **103,** 1648–1656 (1956).
19. Herring, C. Effect of random inhomogeneities on electrical and galvanomagnetic measurements. *J. Appl. Phys.* **31,** 1939–1953 (1960).
20. Sun, Z. G., Mizuguchi, M., Manago, T. & Akinaga, H. Magnetic-field-controllable avalanche breakdown and giant magnetoresistive effects in Gold/semi-insulating GaAs Schottky diode. *Appl. Phys. Lett.* **85,** 5643–5645 (2004).
21. Payling, C. A. *et al.* Electric field-induced quasi-elastic inter-Landau level scattering in the space-charge-limited magnetoconductivity of n+n−n+ InP structures. *Superlatt. Microstruct.* **2,** 415–419 (1986).
22. Yafet, Y., Keyes, R. W. & Adams, E. N. Hydrogen atom in a strong magnetic field. *J. Phys. Chem. Solids* **1,** 137–142 (1956).
23. Sladek, R. J. Magnetically induced impurity banding in n-InSb. *J. Phys. Chem. Solids* **5,** 157–170 (1958).
24. Shklovskii, B. I. & Efros, A. L. *Electronic Properties of Doped Semiconductors* (Springer, 1984).
25. Sampsell, J. B. & Garland, J. C. Current distortion effects and linear magnetoresistance of inclusions in free-electron metals. *Phys. Rev. B* **13,** 583–589 (1976).
26. Stroud, D. & Pan, F. P. Effect of isolated inhomogeneities on the galvanomagnetic properties of solids. *Phys. Rev. B* **13,** 1434–1438 (1976).
27. Bergman, D. J. & Stroud, D. G. High-field magnetotransport in composite conductors: Effective medium approximation. *Phys. Rev. B* **62,** 6603–6613 (2000).
28. Tornow, M., Weiss, D., v., Klitzing, K. & Eberl, K. Anisotropic magnetoresistance of a classical antidot array. *Phys. Rev. Lett.* **77,** 147–150 (1996).
29. Lampert, M. A. & Rose, A. Volume-controlled, two-carrier currents in solids: The injected plasma case. *Phys. Rev.* **121,** 26–37 (1961).
30. Schoonus, J. J. H. M., Bloom, F. L., Wagemans, W., Swagten, H. J. M. & Koopmans, B. Extremely large magnetoresistance in boron-doped silicon. *Phys. Rev. Lett.* **100,** 127202 (2008).

nature

METHODS

Contact geometry and magnetotransport measurement. In this study, the two-terminal resistance measurement was performed for more than 40 devices made from the n-Si and i-Si substrates with various contact geometries. The indium contacts, whose size was typically 0.5 mm × 0.5 mm, were fabricated on the few-millimetre-square (or rectangular) Si substrates (thickness, 525 μm or 380 μm). Good ohmic contacts were obtained by mechanically removing the naturally oxidized layer before making the indium contacts. The distance between the contacts, L, ranged from 0.1 to 5 mm. The present effect is observed regardless of the value of L/W (Fig. 1b). An example device is shown in Fig. 3a. The two-terminal magnetotransport was measured using a source meter (Keithley 6430) in a PPMS-6000 magnetometer (Quantum Design). The magnetic field was applied either perpendicular to the substrate plane or parallel to the current direction. The temperature-dependence measurements were performed from 2 K to about 350 K. In the internal voltage difference and magnetoresistance measurement, a 4.5 mm × 2.8 mm rectangular n-Si substrate (thickness, 380 μm) with current injection and drain electrodes on one side and four voltage probe electrodes on the other side was prepared using soldering and alloying of the indium electrodes on a mechanically etched substrate at 350 °C for five minutes (Fig. 2a). The average width of the electrodes was ~0.3 mm. We measured the voltage difference and magnetoresistance simultaneously at various probe positions, and the two-terminal (injection–drain) magnetoresistance, at 25 K in the PPMS-6000 magnetometer (Quantum Design). The current was injected using a source meter (Keithley 6430), and the voltage drops across the n-Si device were measured using multimeters (Keithley 2000). The magnetic field was applied perpendicular to the substrate plane.

Effect of SiO_2. The silicon substrates used in this study had thermally oxidized (SiO_2) layers of thickness 100 nm. In some samples, the SiO_2 layer was removed by immersing the silicon substrate in a hydrofluoric acid solution for one hour to ensure its perfect removal. The devices that were pre-treated using hydrofluoric acid solution show the same large positive magnetoresistance, which excludes the possibility that the present effect occurs because of some subtle transport mechanism related to the interface states between the silicon and SiO_2.

LETTERS

Friction laws at the nanoscale

Yifei Mo[1], Kevin T. Turner[1,2,3] & Izabela Szlufarska[1,3]

Macroscopic laws of friction do not generally apply to nanoscale contacts. Although continuum mechanics models have been predicted to break down at the nanoscale[1], they continue to be applied for lack of a better theory. An understanding of how friction force depends on applied load and contact area at these scales is essential for the design of miniaturized devices with optimal mechanical performance[2,3]. Here we use large-scale molecular dynamics simulations with realistic force fields to establish friction laws in dry nanoscale contacts. We show that friction force depends linearly on the number of atoms that chemically interact across the contact. By defining the contact area as being proportional to this number of interacting atoms, we show that the macroscopically observed linear relationship between friction force and contact area can be extended to the nanoscale. Our model predicts that as the adhesion between the contacting surfaces is reduced, a transition takes place from nonlinear to linear dependence of friction force on load. This transition is consistent with the results of several nanoscale friction experiments[4–7]. We demonstrate that the breakdown of continuum mechanics can be understood as a result of the rough (multi-asperity) nature of the contact, and show that roughness theories[8–10] of friction can be applied at the nanoscale.

According to the well-known Amontons' laws[11], formulated in 1699, the friction force F_f between two macroscopic bodies is linearly proportional to the applied load L, that is, $F_f = \bar{\mu}L$, where $\bar{\mu}$ is the macroscopic coefficient of friction. F_f is also independent of the macroscopic contact area A_{macro}. It was later noted[12] that a macroscopic contact is rough and consists of a large number of smaller contacts (so-called asperities), whose total area $\sum A_{asp}$ is much smaller than A_{macro}. The friction force was shown to be proportional to this true contact area, that is, $F_f = \bar{\tau} \sum A_{asp}$, where $\bar{\tau}$ is an effective shear strength of the contacting bodies (Table 1). To isolate surface roughness from other physical parameters that affect friction, many experiments and simulations have focused on studies of individual asperities, which have radii of curvature from tens of nanometres to micrometres in size and are assumed to be perfectly smooth[13–15].

Concurrently with single-asperity studies, roughness contact theories are being developed[8–10,16] to bridge the gap between the mechanics of single asperities and that of macroscopic contacts.

One reason single-asperity measurements have been so successful is that deformation of an asperity can be described by continuum mechanics theories (Table 1). A model for non-adhesive contact between homogenous, isotropic, linear elastic spheres, was first developed by Hertz[17], who showed that $A_{asp} \propto L^{2/3}$. Adhesion effects were included in a number of subsequent models, among which Maugis-Dugdale theory[18] has been frequently used because of its high degree of flexibility. The common feature of all the single-asperity theories is that A_{asp} is a sublinear function of L. In addition, several scanning force microscopy (SFM) studies have reported that $F_f = \tau A_{asp}$, with τ being an interfacial shear strength[13,19]. In this case, F_f will be a sublinear function of load[20].

A key question is whether single-asperity models can describe contact behaviour at the nanoscale. Deviations from single-asperity theories have sometimes been observed at these length scales and they have been attributed to the break-down of continuum mechanics. For instance, SFM experiments have been reported[5,6,21] where $F_f \propto L$ or even where these two quantities are nearly independent of each other[22]. More detailed discussion of the evidence for the break-down of continuum mechanics at the nanoscale can be found in ref. 13. The challenge in using SFM to establish friction laws at the nanoscale lies in the large sensitivity of contact behaviour to specific experimental conditions, such as surface chemistry or tip geometry. These challenges can be overcome in molecular dynamics simulations[1,5,20,23,24], which have been invaluable in identifying atomistic phenomena underlying friction. However, the use of idealized tip models or force fields in such molecular dynamics studies has led to contradictory results regarding friction laws. For instance, Luan and Robbins[1] and Wenning and Müser[25] found that for incommensurate non-adhesive contacts $F_f \propto L^{2/3}$, which was in contrast to later simulations by Gao et al.[20] who found $F_f \propto L$.

Here we perform large-scale molecular dynamics simulations of SFM experiments to determine friction laws at the nanoscale. We simulate

Table 1 | Summary of friction laws

Friction laws	F_f versus area	F_f versus L	Notes
Macroscale theories			
Amontons' law	Independent of A_{macro}	$F_f = \bar{\mu} \cdot L$	Law first discovered by Leonardo da Vinci
Bowden and Tabor	$F_f = \bar{\tau} \cdot \sum A_{asp}$	$F_f = \bar{\mu} \cdot L$	Law results from contact roughness
Single-asperity theories			
Non-adhesive (based on Hertz model)	$F_f = \tau \cdot A_{asp}$	$F_f \propto L^{2/3}$	Linear dependence of F_f on A_{asp} is generally believed to be true for microscale contacts, but has been questioned for nanoscale contacts
Adhesive (for example, Maugis–Dugdale)	$F_f = \tau \cdot A_{asp}$	Sublinear	
Multi-asperity picture of nanoscale contact (our model)			
Non-adhesive	$F_f = \tau \cdot A_{real} = \tau \cdot N_{at} \cdot A_{at}$ $F_f \neq \tau \cdot A_{asp}$	$F_f = \mu \cdot L$	Dependence of F_f on A_{real} has been directly verified. Linear dependence of F_f on L is due to atomic roughness and small contact pressures
Adhesive	$F_f = \tau \cdot A_{real} = \tau \cdot N_{at} \cdot A_{at}$ $F_f \neq \tau \cdot A_{asp}$	Sublinear	Adhesion induces transition from linear to sublinear behaviour

A_{macro} is the macroscopic contact area. A_{asp} is the contact area of a single asperity; A_{real} is the real contact area defined as the number of atoms N_{at} in contact multiplied by the average contact area A_{at} of an interfacial atom.

sliding under a normal load and we determine the dependence of friction force on the applied load, contact area, and on the presence of adhesion at the interface. We focus on amorphous carbon tips and diamond samples (Fig. 1a), both terminated with hydrogen, because
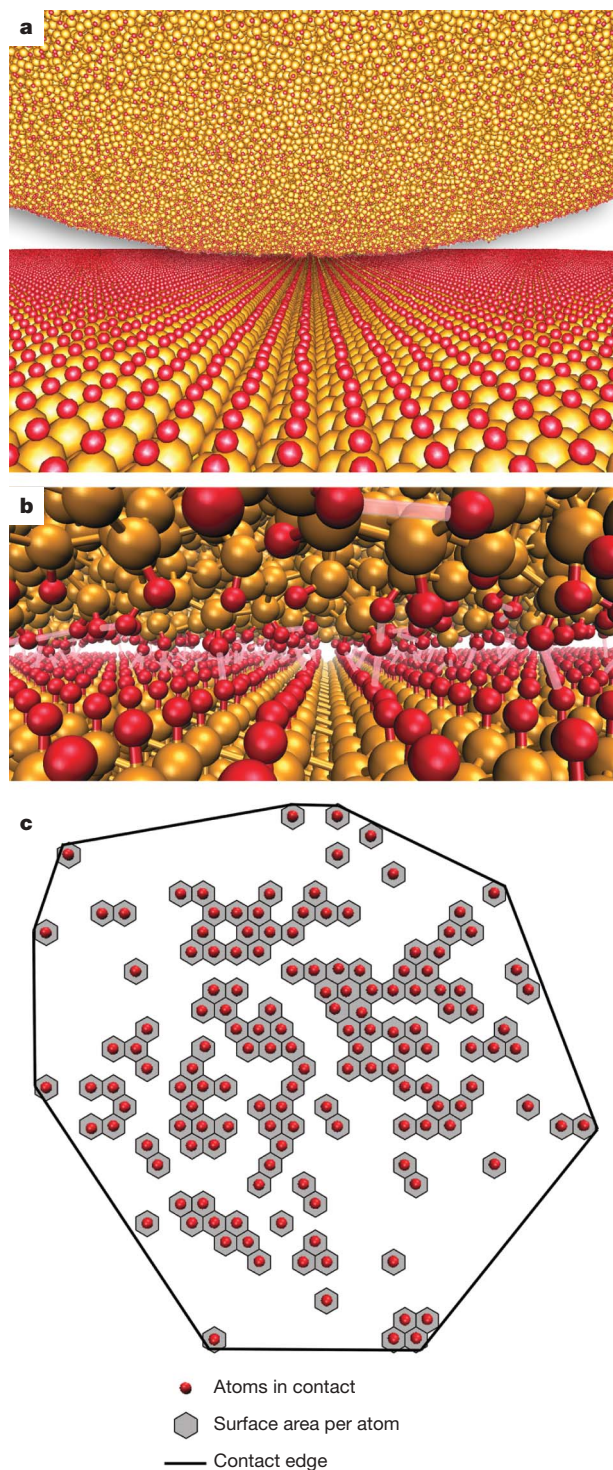

Figure 1 | **Contact between an amorphous carbon tip and a diamond sample. a,** Far view, showing contact geometry. Golden and red atoms correspond to C and H, respectively. **b,** Close view. Solid red and golden sticks represent covalent bonds. Translucent pink sticks represent repulsive interactions. **c,** Contact area definitions. Red circles represent sample atoms within the range of chemical interactions from tip atoms. Contact area per atom A_{at} is represented by grey hexagons. Real contact area A_{real} is the sum of the areas of hexagons. The contact area A_{asp} of an asperity is enclosed by the edge (solid line) of the contact zone.

of the availability of experimental data[20] that can be used for comparison. We use realistic force fields[26], spherical tips that are allowed to deform together with the sample and that have curvature radii R of up to 30 nm (see Supplementary Fig. S1), and the simulations are performed at the temperature of 300 K. The force fields consist of short-range chemical interactions and long-range van der Waals interactions.

Defining contact area is one of the major challenges for understanding friction in nanoscale contacts. Because fundamentally contact is formed by atoms interacting across the interface (Fig. 1b), we define a real contact area to be $A_{real} = N_{at}A_{at}$, where N_{at} is the number of atoms of the sample within the range of chemical interactions from the tip atoms and A_{at} is the average surface area per atom. In continuum mechanics A_{asp} is defined by the edge of the contact zone. The definition of the contact edge becomes ambiguous when the atomistic nature of the interface dominates its physical behaviour[27,28]. Here, we define A_{asp} to be the area enclosed by a convex hull around atoms in contact, as shown in Fig. 1c.

We first perform molecular dynamics simulations of normal loading and friction in the absence of van der Waals forces. Because both the tip and the sample are passivated with hydrogen atoms, adhesion due to short-range chemical forces is negligible. In this case a non-adhesive single asperity (Hertz) model is expected to apply, that is, $A_{asp} = \pi(3R/4E^*)^{2/3}L^{2/3}$, where $E^* = [(1-v_1^2)/E_1 + (1-v_2^2)/E_2]^{-1}$ is the effective modulus of the contact, E_1 and E_2 are the tip and the sample Young's moduli, and v_1 and v_2 are the tip and the sample Poisson's ratios, respectively. With the additional assumption of $F_f = \tau A_{asp}$, the Hertz model predicts that $F_f \propto L^{2/3}$. Our calculations show that $A_{asp} \propto L^{0.7}$ and $F_f \propto L$ (Fig. 2a), which yields $F_f \neq \tau A_{asp}$. Additionally, even though A_{asp} obtained in simulations shows approximately the same 2/3 power-law dependence on L as in the Hertz theory, the effective modulus E^* values calculated from fitting the Hertz model to the simulation data are 61% smaller than the value calculated directly from the definition of E^* (see Supplementary Methods). Our results confirm the conclusions of other authors that single-asperity theories break down at the nanoscale[1,5]. To account for the fact that F_f is not linear with A_{asp}, Wenning and Müser[25] suggested that τ is not a constant, but varies with contact pressure. Other authors proposed an empirical model in which mechanics of a nanoscale non-adhesive contact is controlled by load, that is, $F_f = \mu L$ and the contact area is undefined and unnecessary[5,29].

We argue that the break-down of single-asperity theories of friction is due to the fact that at these length scales the real contact area A_{real} is different from A_{asp} (see Fig. 1c) and that friction laws should be defined in terms of A_{real}. Our simulations show that $A_{real} \propto L$ (Fig. 2b) and $F_f = \tau A_{real}$ with constant τ (Fig. 2c), which is consistent with the relation $F_f \propto L$ (Fig. 2a). As shown in Table 1, friction force is now proportional to contact area at all length scales as long as the contact area is correctly defined at each length scale. Also, realizing that it is A_{real} (or, more fundamentally, N_{at}) that controls friction at the nanoscale, we can now understand why $F_f \propto L$. Such linear dependence is characteristic of rough contacts (Table 1) and in our case is a result of atomic roughness, as shown by the fact that $A_{real} \neq A_{asp}$. The above results demonstrate that a nanoscale contact, which had been previously viewed as a single entity, consists of yet smaller contacts of atomic size. Macroscale roughness theories can be applied to describe the behaviour of nanoscale contacts (see Supplementary Methods).

We investigate the effect of van der Waals adhesion on contact behaviour by adding these forces to the tip–sample interactions and by performing additional molecular dynamics simulations. As shown in Fig. 3a, the relation $F_f = \tau A_{real}$ still holds, which demonstrates that friction is controlled by the short-range (chemical) interactions even in the presence of dispersive forces. However, unlike the non-adhesive case, here F_f is a sublinear function of L (Fig. 3b), which is consistent with predictions of adhesive single-asperity models (see Table 1). We fit $F_f(L)$ to the Maugis–Dugdale model using a convenient approximation proposed by Carpick, Ogletree, and Salmeron (COS)[30] and later physically justified by Schwarz[31]. The fits show an excellent agreement with both
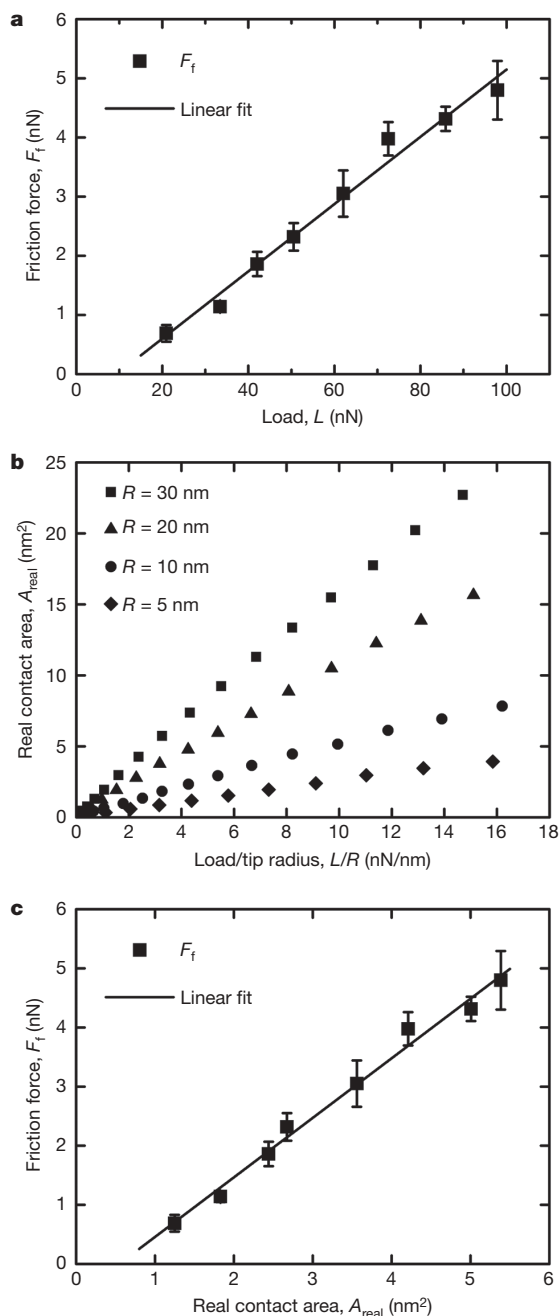
Figure 2 | Mechanics of non-adhesive contacts. a, Friction force versus load. **b**, Real contact area versus load. **c**, Friction force versus contact area. F_f is measured over five lattice periods of the sample's surface. The error bars in **a** and **c** correspond to a standard deviation in these measurements. Linear extrapolation to $F_f \rightarrow 0$ leads to finite values of A_{real} and L, but it is not clear whether such extrapolation should be linear. Measurements at very small loads carry large uncertainty owing to the small number of atoms in contact and to thermal vibrations.

Figure 3 | Mechanics of adhesive contacts. a, Friction force versus real contact area. **b**, Friction force versus load. **c**, Contributions to the total load as a function of real contact area: total load L (full squares); van der Waals contribution L_{vdW} (empty triangles); elastic restoring force $L_{el} = L - L_{vdW}$ (empty circles). F_f is measured over three lattice periods of the sample's surface. The error bars in **a** and **b** correspond to a standard deviation in these measurements.

the Maugis–Dugdale theory and with experimental results (see Methods Summary), which validates our modelling approach. For instance, the fitted value of Tabor's[32] parameter μ_T is 0.19, which falls within the range of 0.11–0.22 calculated directly from theory for our system.

However, the fact that adhesive single-asperity theory fits both our simulation data and experimental data very well should not be taken as evidence that the theory represents the correct physics. We argue that the flexibility of the Maugis–Dugdale model (that is, three fitting parameters) masks the model's physical deficiencies. For instance, we can check whether the elastic restoring force L_{el}, calculated by subtracting van der Waals forces L_{vdW} from the total load L, follows Hertz's predictions,

that is, whether $A_{real} \propto L_{el}^{2/3}$. As can be seen from Fig. 3c, $A_{real} \propto L_{el}$, which means that the correct physical picture is that of a rough (multi-asperity) contact rather than of a single-asperity one (see Table 1).

Why then does F_f depend sublinearly on L? This behaviour is due to the presence of adhesion forces L_{vdW}, which for a spherical tip in contact with a flat sample do not scale linearly with A_{real} (triangles in Fig. 3c). As a result the total load $L = L_{el} + L_{vdW}$ is not proportional to A_{real} and hence $F_f = \tau A_{real}$ is a sublinear function of L. Our models therefore predict that as adhesion in the contact is increased, a transition takes place from linear to sublinear dependence of F_f on L. The opposite effect can be achieved by increasing the roughness of the

interface. Behaviour consistent with single-asperity theories can be expected in the limit of $A_{real} \approx A_{asp}$, for example, for large contact pressures and for tip radii in the micrometre range and larger[9,33].

The identification of the multi-asperity nature of dry nanoscale contacts and of the associated laws of friction allows controversies that exist in the field to be resolved. For instance, the sublinear dependence of friction on load for non-adhesive incommensurate contacts reported by Luan and Robbins[1] is probably due to the relatively low roughness of the simulated interfaces. Two effects could contribute to this low surface roughness. First, the simulated surfaces were bare, that is, they were not passivated with adsorbate atoms. Second, the repulsive part of the Lennard–Jones force field that was used in the simulations is physically too stiff, which reduces corrugation of the surface potential energy. Low surface corrugation (roughness) results in $A_{real} \approx A_{asp}$, which explains why the authors observed a sublinear dependence of F_f on L. Gao et al.[20] simulated the opposite limit of contact behaviour. These authors used more realistic force fields and they simulated surfaces passivated with hydrogen atoms, which yielded atomically rough interfaces and resulted in the $F_f \propto L$ dependence. Atomic roughness was additionally frozen in by making the tip rigid. The transition from sublinear to linear $F_f(L)$ dependence due to adhesion reduction has been previously observed experimentally[4–6,34,35]. We can now understand that such transition takes place when contact roughness, defined by the corrugation of surface potential energy, becomes large in comparison to the range of interfacial interactions. Adhesion reduction can be realized for instance by surface passivation[6] or surface functionalization[4,34]. The transition from sublinear to linear friction can occur also during sliding if the surfaces are damaged (that is, roughened) in the process[5].

METHODS SUMMARY

In molecular dynamics simulations we use the reactive empirical bond-order (REBO) potential[26]. The REBO potential accurately describes the cohesive energies and chemical reactions of hydrocarbon systems as well as the elastic constants of solid carbon-based materials. The range of the REBO potential extends as far as the chemical interactions and does not include dispersive forces. The van der Waals interactions are therefore integrated with REBO using an analytical switching function in the regime where the two potentials overlap. For detailed description of the simulation techniques and schedule see Supplementary Methods.

We have also performed finite element analysis to demonstrate that the effect of boundary conditions in our simulations is negligible, that is, if our tips were represented by continuous, linear-elastic solids with smooth interfaces, the contact behaviour would be accurately described by continuum single-asperity models.

Our simulations show an excellent agreement with theory and experiments. For example, the F_f versus L dependence obtained in our simulations for an $R = 30$ nm adhesive tip sliding over H terminated diamond (111) surface was fitted to the Maugis–Dugdale model (see Supplementary Methods). The fitted COS transition parameter is 0.21, which yields the Tabor's[32] parameter μ_T of 0.19. Direct calculation of μ_T, based on our system geometry and elastic properties, gives a range from 0.11 to 0.22, which brackets the fitted value of μ_T. Our fits show that μ_T decreases with decreasing R, which is also in agreement with single-asperity models. Both the shear strength τ and the contact pressure are of the same order of magnitude as in SFM experiments[20]. Specifically, using a procedure similar to that in ref. 20 for $R = 30$ nm we estimate that $\tau \approx 508$ MPa and the contact pressure at $L \approx 100$ nN is 6.12 GPa.

1.   Luan, B. & Robbins, M. O. The breakdown of continuum models for mechanical contacts. *Nature* 435, 929–932 (2005).
2.   Delrio, F. W. et al. The role of van der Waals forces in adhesion of micromachined surfaces. *Nature Mater.* 4, 629–634 (2005).
3.   Zykova-Timan, T., Ceresoli, D. & Tosatti, E. Peak effect versus skating in high-temperature nanofriction. *Nature Mater.* 6, 230–234 (2007).
4.   Colburn, T. J. & Leggett, G. J. Influence of solvent environment and tip chemistry on the contact mechanics of tip-sample interactions in friction force microscopy of self-assembled monolayers of mercaptoundecanoic acid and dodecanethiol. *Langmuir* 23, 4959–4964 (2007).
5.   Gao, J. et al. Frictional forces and Amontons' law: from the molecular to the macroscopic scale. *J. Phys. Chem.* 108, 3410–3425 (2004).
6.   Grierson, D. S. *Nanotribological Properties of Nanostructured Hard Carbon Thin Films.* PhD thesis, Univ. Wisconsin (2008).
7.   Ruths, M. Boundary friction of aromatic self-assembled monolayers: Comparison of systems with one or both sliding surfaces covered with a thiol monolayer. *Langmuir* 19, 6788–6795 (2003).
8.   Müser, M. H. A rigorous, field-theoretical approach to the contact mechanics of rough, elastic solids. *Phys. Rev. Lett.* 100, 055504 (2008).
9.   Persson, B. N. J. Theory of rubber friction and contact mechanics. *J. Chem. Phys.* 115, 3840–3861 (2001).
10. Greenwood, J. A. & Williamson, J. B. P. Contact of nominally flat surfaces. *Proc. R. Soc. Lond. A* 295, 300–319 (1966).
11. Amontons, G. De la resistance causée dans les machines. *Mem. Acad. R. A* 275–282 (1699).
12. Bowden, F. P. & Tabor, D. *The Friction and Lubrication of Solids* (Clarendon, 1950).
13. Szlufarska, I., Chandross, M. & Carpick, R. W. Recent advances in single-asperity nanotribology. *J. Phys. D* 41, 123001 (2008).
14. Binnig, G., Quate, C. F. & Gerber, C. Atomic force microscope. *Phys. Rev. Lett.* 56, 930–933 (1986).
15. Perry, S. S. Scanning probe microscopy measurements of friction. *MRS Bull.* 29, 478–483 (2004).
16. Bush, A. W., Gibson, R. D. & Thomas, T. R. The elastic contact of a rough surface. *Wear* 35, 87–111 (1975).
17. Hertz, H. On the contact of elastic solids. *J. Reine Angew. Math.* 92, 156 (1881).
18. Maugis, D. Adhesion of spheres—the JKR-DMT transition using a Dugdale model. *J. Colloid Interface Sci.* 150, 243–269 (1992).
19. Riedo, E., Chevrier, J., Comin, F. & Brune, H. Nanotribology of carbon based thin films: The influence of film structure and surface morphology. *Surf. Sci.* 477, 25–34 (2001).
20. Gao, G. T., Cannara, R. J., Carpick, R. W. & Harrison, J. A. Atomic-scale friction on diamond: a comparison of different sliding directions on (001) and (111) surfaces using MD and AFM. *Langmuir* 23, 5394–5405 (2007).
21. Pietrement, O. & Troyon, M. Study of the interfacial shear strength pressure dependence by modulated lateral force microscopy. *Langmuir* 17, 6540–6546 (2001).
22. Socoliuc, A., Bennewitz, R., Gnecco, E. & Meyer, E. Transition from stick-slip to continuous sliding in atomic friction: Engineering a new regime of ultralow friction. *Phys. Rev. Lett.* 92, 134301 (2004).
23. Chandross, M., Lorenz, C. D., Stevens, M. & Grest, G. S. Simulations of nanotribology with realistic probe tip models. *Langmuir* 24, 1240–1246 (2008).
24. Szlufarska, I., Nakano, A. & Vashishta, P. A crossover in the mechanical response of nanocrystalline ceramics. *Science* 309, 911–914 (2005).
25. Wenning, L. & Müser, M. H. Friction laws for elastic nanoscale contacts. *Europhys. Lett.* 54, 693–699 (2001).
26. Brenner, D. et al. A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons. *J. Phys. Condens. Matter* 14, 783–802 (2002).
27. Johnson, K. L. Adhesion and friction between a smooth elastic spherical asperity and a plane surface. *Proc. R. Soc. Lond. Ser. A* 453, 163–179 (1997).
28. Luan, B. & Robbins, M. O. Contact of single asperities with varying adhesion: comparing continuum mechanics to atomistic simulations. *Phys. Rev. E* 74, 026111 (2006).
29. Israelachvili, J. N. & Berman, A. D. in *Handbook of Micro/Nanotribology* (ed. Bhushan, B.) 2nd edn, 371–432 (CRC Press, 1999).
30. Carpick, R. W., Ogletree, D. F. & Salmeron, M. A general equation for fitting contact area and friction vs load measurements. *J. Colloid Interface Sci.* 211, 395–400 (1999).
31. Schwarz, U. D. A generalized analytical model for the elastic deformation of an adhesive contact between a sphere and a flat surface. *J. Colloid Interface Sci.* 261, 99–106 (2003).
32. Tabor, D. Surface forces and surface interactions. *J. Colloid Interface Sci.* 58, 2–13 (1977).
33. Horn, R. G., Israelachvili, J. N. & Pribac, F. Measurement of the deformation and adhesion of solids in contact. *J. Colloid Interface Sci.* 115, 480–492 (1987).
34. Ruths, M., Alcantar, N. A. & Israelachvili, J. N. Boundary friction of aromatic silane self-assembled monolayers measured with the surface forces apparatus and friction force microscopy. *J. Phys. Chem. B* 107, 11149–11157 (2003).
35. Putman, C. A. J., Igarashi, V. & Kaneko, R. Single-asperity friction in friction force microscopy — The composite-tip model. *Appl. Phys. Lett.* 66, 3221–3223 (1995).

Author Contributions I.S. conceived the simulations and Y.M carried them out. Y.M. and I.S. designed the data analysis and Y.M. carried it out. I.S. and Y.M. co-wrote the paper. K.T.T. provided guidance and software for finite element analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to I.S. (izabela@engr.wisc.edu).

LETTERS

Giant aeolian dune size determined by the average depth of the atmospheric boundary layer

Bruno Andreotti[1], Antoine Fourrière[1], Fouzia Ould-Kaddour[2], Brad Murray[3] & Philippe Claudin[1]

Depending on the wind regime[1,2], sand dunes exhibit linear[3,4], crescent-shaped or star-like forms[5] resulting from the interaction between dune morphology and sand transport[6–8]. Small-scale dunes form by destabilization of the sand bed[9–11] with a wavelength (a few tens of metres) determined by the sand transport saturation length[11–13]. The mechanisms controlling the formation of giant dunes, and in particular accounting for their typical time and length scales, have remained unknown. Using a combination of field measurements and aerodynamic calculations, we show here that the growth of aeolian giant dunes, ascribed to the non-linear interaction between small-scale superimposed dunes[4,10,14,15], is limited by the confinement of the flow within the atmospheric boundary layer[16,17]. Aeolian giant dunes and river dunes form by similar processes, with the thermal inversion layer that caps the convective boundary layer in the atmosphere[18] acting analogously to the water surface in rivers. In both cases, the bed topography excites surface waves on the interface that in turn modify the near-bed flow velocity. This mechanism is a stabilizing process that prevents the scale of the pattern from coarsening beyond the resonant condition. Our results can explain the mean spacing of aeolian giant dunes ranging from 300 m in coastal terrestrial deserts to 3.5 km. We propose that our findings could serve as a starting point for the modelling of long-term evolution of desert landscapes under specific wind regimes.

Aeolian dune fields generically present two well-separated length scales[15,19,20] (Fig. 1). The smallest superimposed bed-forms have been explained by a linear aerodynamic instability the initial wavelength λ_s of which is related to the length needed for sand transport to reach equilibrium with the wind strength[11–13] and are thus called 'elementary dunes'. Tentative explanations for the formation of giant dunes have proposed specific dynamics associated with the different types of dunes, thought of as isolated objects[4,5]. In contrast, we hypothesize that the emergence of giant transverse, longitudinal and star dunes results from collective processes, the symmetries of the different patterns resulting from those of the wind rose. Indeed, both small- and large-scale dunes share the same symmetries under a common wind regime (Fig. 1). Furthermore, the different types of giant dunes exhibit the same characteristic wavelength λ_g, of the order of a kilometre (Fig. 1), and feature superimposed structures the avalanche slip faces of which show separations on the same scale λ_s. These commonalities suggest that the key dynamical processes forming giant dunes are not to be looked for in the shape of the pattern but in the understanding of the characteristic time and length scales resulting from their formation. We propose a novel collective mechanism in which the average structure of the atmosphere determines the giant scale.

The dryness of deserts results primarily from the overall stability of the atmosphere in anticyclonic regions. The stable stratification of the free atmosphere, characterized by the Brunt-Väisälä frequency $N = \sqrt{\frac{g}{\theta}\frac{d\theta}{dz}}$ (where g is the gravity and θ is the virtual potential temperature) for restoring vertically displaced air parcels, prevents the development of turbulence. The temperature gradient $\gamma = d\theta/dz \approx 4\,\mathrm{K\,km^{-1}}$ is largely independent of the location and the season (Supplementary Information 1). In winter, the heat flux from the surface is insufficient to produce convection, so the atmospheric boundary layer is stably stratified almost down to the ground (Fig. 2b). In warmer seasons, however, a convective (well-mixed) boundary layer forms, in which the temperature and the horizontal wind velocity are roughly homogeneous. A thermal inversion (capping) layer, characterized by a jump in air density $\Delta\rho$ separates stable and unstable layers. The shear stress τ decreases linearly with height in the well-mixed layer, vanishes at the capping altitude H and remains null in the free atmosphere[18]. The base flow is thus similar to that in a river. Furthermore, gravity waves can propagate at a speed c on this atmospheric interface, as they do along the free surface of a river[21]. Introducing the wavenumber $k = 2\pi/\lambda$, the surface-wave dispersion relation can be approximated by:

$$c^2 + \left(1 - \frac{\Delta\rho}{\rho_0}\right)c\sqrt{c^2 - \frac{N^2}{k^2}} = \frac{\Delta\rho}{\rho_0}\frac{g}{k} \qquad (1)$$

in the limit of large kH. The river case is recovered for $\Delta\rho \approx \rho_0$ and $N = 0$. Consequently, aeolian giant dune relief can excite standing surface waves that in turn modify the wind speed and thus the sand transport.

Because the temperature profile $\theta(z)$ in the free atmosphere does not depend much on the season, the typical height of the capping layer H increases linearly with the ground temperature annual variations: $H \approx \delta\theta/\gamma$ (Fig. 2b). The ground temperature is controlled by thermal transfers at the synoptic scale and varies annually from a few degrees in temperate environments close to the seashore to a few tens of degrees in more continental regions (Supplementary Information 1). We expect modern temperature measurements to be relevant over the current Holocene interglacial. We have measured the giant dune wavelength λ in all homogeneous desert regions of the world (Supplementary Information 2) and shown that it is linearly correlated with $\delta\theta/\gamma$, independent of the dune symmetry (Fig. 2a). These data then strongly support the idea that the size of giant dunes scales with the depth of the atmospheric boundary layer.

To understand the origin of this relation, we have computed the two-dimensional turbulent flow confined between a capping stratified fluid and a wavy bottom (Supplementary Information 7). Navier-Stokes equations are closed by the means of a Prandtl mixing length approach, and are expanded with respect to the bottom corrugation aspect ratio a/λ, which is a small parameter. A key output of these calculations is the phase shift between the bottom topography and the

---

[1]Laboratoire de Physique et Mécanique des Milieux Hétérogènes (PMMH UMR 7636 CNRS-ESPCI-P6-P7), 10 rue Vauquelin, 75005 Paris, France. [2]Laboratoire de Physique Théorique, Université Abou Bekr Belkaid, Tlemcen, Algeria. [3]Nicholas School of the Environment and Earth Sciences, Center for Nonlinear and Complex Systems, Duke University, Box 90230, Durham, North Carolina 27708-0230, USA.
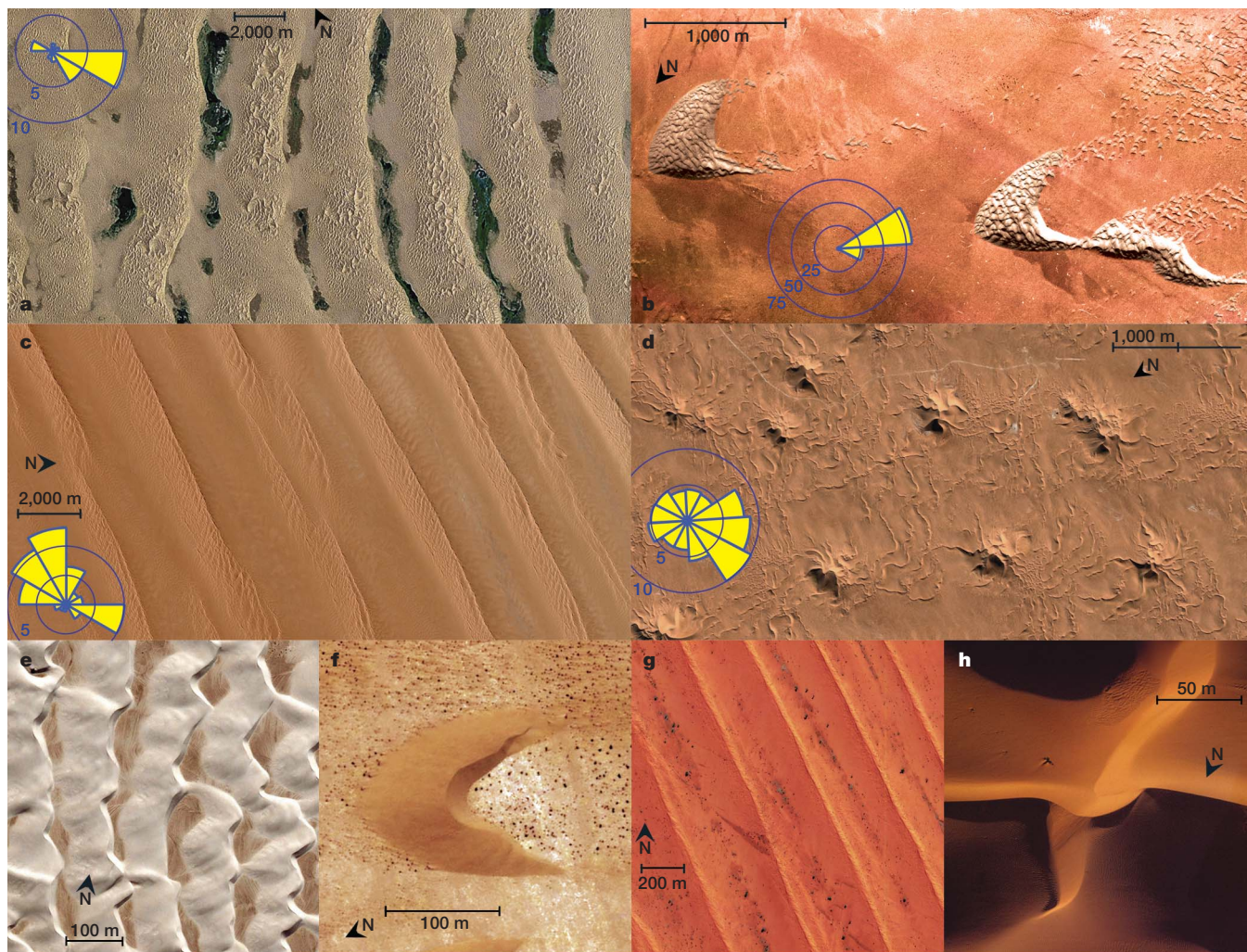
Figure 1 | Separation of length scales between small and giant dunes for the different wind regimes. a–d, Giant dunes. **a,** Transverse dunes, Badain Jaran (China, 38° 38′ N/104° 59′ E); **b,** crescent-shaped barchan, Atlantic Sahara (Morocco, 28° 02′ N/12° 11′ E); **c,** longitudinal dunes, RubAlKhali, (South Arabia, 18° 11′ N/47° 21′ E); **d,** star dunes, Grand Erg Oriental (Algeria, 31° 27′ N/07° 45′ E). **e–h,** Small dunes. **e,** White sands (USA, 32° 49′ N/106° 16′ W); **f,** Atlantic Sahara (Morocco, 27° 11′ N/13° 13′ W);

g, Australia (23° 51′ S/136° 33′ E); **h,** Mauritania (18° 09′ N/15° 29′ W). These aerial photographs (credits: Digital Globe) show that giant dunes are not isolated objects but involve a composite spatial organization. The wind regimes are characterized by the averaged sand flux roses (1999–2007), on which yellow arrows indicate the direction in which the dominant wind moves. The symmetries of these roses are consistent with those of the giant dune pattern (Supplementary Information 5).
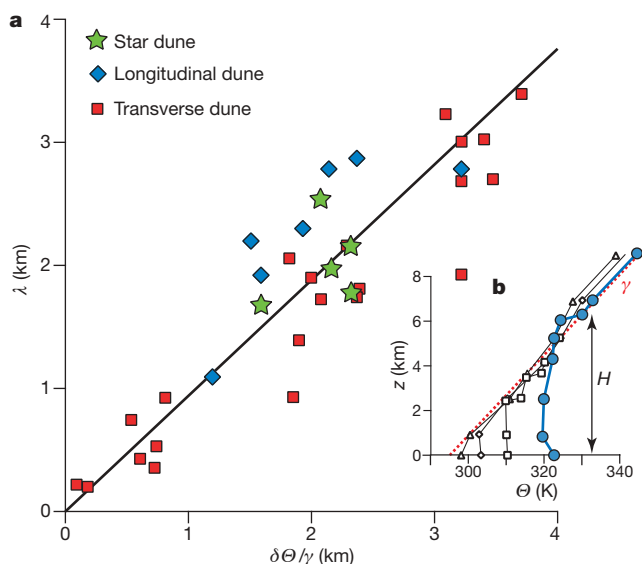
basal shear stress τ_b, which controls the linear growth of dunes[9–11]. Figure 3 shows that the fastest-growing wavelength always corresponds to elementary dunes ($\lambda = \lambda_s$). Furthermore, the presence of

Figure 2 | Selection of the wavelength of giant dunes by the depth of the atmospheric boundary layer. a, Measured giant dune wavelength λ as a function of the characteristic mixing height, computed from the annual ground temperature variations $\delta\theta$ (Supplementary Table). The solid line is the identity: $\lambda = \delta\theta/\gamma$. λ in some cases is the average dune spacing and in others is the peak wavelength in a Fourier spectrum (Supplementary Information 2). As the dune patterns present dislocation defects, the distribution of dune spacing must have a width of the order of λ. $\delta\theta$ is averaged over the period 2000–2007 (Supplementary Information 1). $\delta\theta$ mostly reflects the distance to the ocean, and is therefore likely to remain approximately constant at any location over the timescale of formation of giant dunes. **b,** Virtual potential temperature profiles $\theta(z)$, measured by sounding balloons at noon, in Sulayel (South Arabia, 20° 47′ N/45° 67′ E), at different seasons (blue circles: 27/07/1978, squares: 22/10/1978, diamonds: 01/12/1977, triangles: 12/02/1978). The free atmosphere preserves an almost steady linear profile (red dashed line) of slope $\gamma = 4$ K km^{-1}. The summer profile shows a 6-km-deep mixed layer. The winter profile is stably stratified nearly down to the soil. The residual mixing height (<100 m) is the altitude at which the mechanically produced turbulence is in balance with the dissipative effect of negative buoyancy.

the capping layer has an overall stabilizing effect, which is particularly pronounced at the scale H. When the fluid flows at the velocity U over a relief of wavelength λ, waves on the capping layer are generated at this wavelength. The system is resonant when U equals the wave propagation speed c at the wavenumber k. This situation corresponds to a shift of the maximum basal shear stress to downwind of the crest (Fig. 3), and thus to a stable situation in which bed perturbations decay[9–11]. With a free surface, the growth rate σ for wavelengths of the order of H is lower than it would be without a free surface (Fig. 3). Thus, the formation of giant dunes cannot be explained by a primary linear instability, that is, by a new growth-rate peak caused by the free surface.

To go beyond linear aerodynamics, we have extended the expansion of the flow field up to the third order in the aspect ratio a/λ, to get the first nonlinear corrections to the basal shear stress (Supplementary Information 7). The phase shift with respect to the bed eventually vanishes for some amplitude, which can be interpreted as the equilibrium height of a mature dune of a certain wavelength. For small λ, the selected aspect ratio is roughly constant ($a/\lambda \approx 1/12$) and matches quantitative observations[22,23] (Fig. 4c). For larger wavelengths, the presence of the capping layer causes dunes to stabilize at smaller aspect ratios (Fig. 4c). The resulting relation between the steady amplitude a and the wavelength λ presents a pronounced maximum, not very
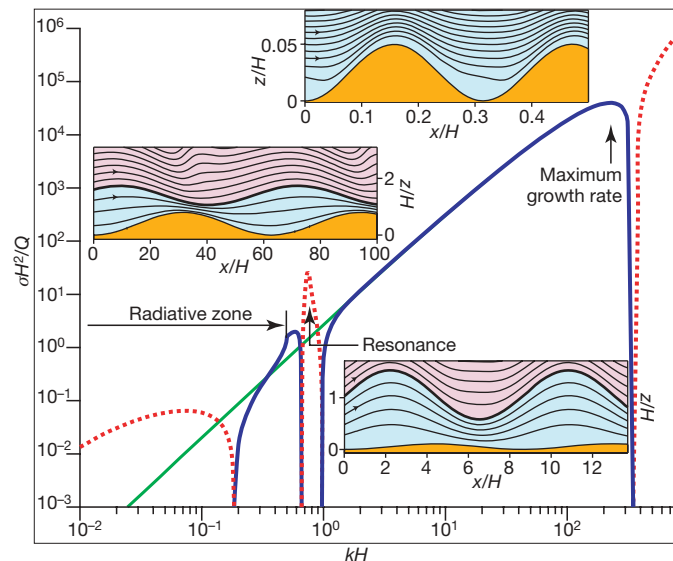
Figure 3 | Linear stability analysis of the dune formation process. Growth rate modulus $|\sigma|$ rescaled by the timescale H^2/Q (Q is the reference saturated sand flux[10]; H^2/Q ranges from 2,500 years in the Atlantic Sahara to 250,000 years in RubAlKhali) as a function of the rescaled wavenumber kH. Stable wavelengths ($\sigma < 0$) are shown by a dotted red line and unstable ones ($\sigma > 0$) are shown by a solid blue line. The curve obtained in the absence of a capping layer (in the limit of infinite H) is shown in green. Insets show streamlines over sinusoidal dunes (orange) computed at the linear order. The stratification of the free atmosphere (purple) is characterized by the non-dimensional number $B = U/HN = 2$, and the density discontinuity across the capping layer (bold line) that separates the well-mixed layer (blue) from the free atmosphere by $S = \rho_0 U^2/\Delta\rho gH = 0.7$. The Reynolds number is $Re = U\lambda/v = 10^8$. The wind blows from left to right (shown by arrows). The area where near-bed streamlines are most dense corresponds to the maximum shear stress, which separates an erosion zone upstream from a zone of accretion. Thus, if the maximum occurs upstream of the crest of a bed perturbation, the bump grows, and vice versa. The growth rate of small wavelength ($\lambda \ll H$) scales as Qk^2 and is maximum at λ_s. Accordingly, the streamlines are squeezed upstream of the dune crest (top inset, computed for $kH = 20$). At the surface wave resonance, for $U = c(k)$, the capping layer deforms such that the streamlines are squeezed downstream of the relief crest (bottom inset, computed for $kH = 0.7$). Beyond the radiative threshold, for $Uk < N$, standing internal waves are generated in the free atmosphere, which leads to the same stabilizing effect (left inset, computed for $kH = 0.1$).
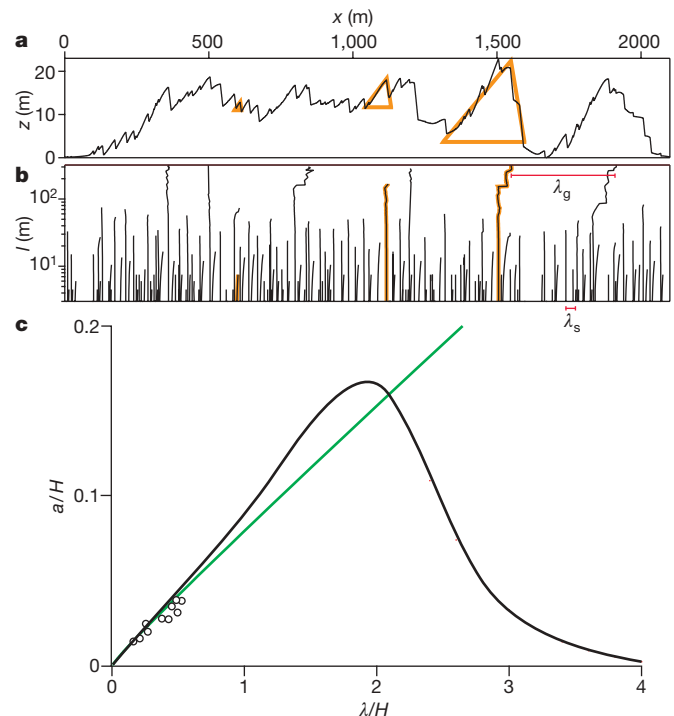
Figure 4 | Pattern coarsening and nonlinear wavelength selection. **a**, Profile of transverse giant dunes in the Atlantic Sahara (28° 03′ N/12° 12′ W). The combination of GPS data and direct geometrical measurements (lengths, heights and angles) ensures an excellent relative and absolute accuracy. Nascent superimposed dunes are detected down to centimetre-scale amplitudes. **b**, Wavelet transform maxima represented in a scale-space diagram. In this multi-scale decomposition, each vertical line indicates the presence of a triangle-like sub-structure (see orange ones for example), whose length and position are given by the upper tip coordinates. **c**, Amplitude a of steady dunes as a function of their wavelength λ (both rescaled by H), predicted by the nonlinear aerodynamic model with a capping layer (black line, $B = U/HN = 0.7$, $S = \rho_0 U^2/\Delta\rho gH = 0.7$ and $Re = U\lambda/v = 10^8$) and without a capping layer (green line). Circles correspond to field data[22,23]. Confining the flow below the capping layer tends to compress streamlines over dune crests, reducing the steady amplitude (compared to the case without the capping layer) as wavelength increases from small values. Free-surface-wave effects then cause steady amplitudes to approach zero as wavelengths increase further.

sensitive to variations in parameters (Fig. 4c), and which corresponds to giant dunes commensurate with H.

As is supported by Ground Penetrating Radar studies of Pleistocene giant dunes[14], giant dunes are likely to result from the amalgamation[15] of superimposed dunes. Such progressive growth, ultimately limited by the average depth of the atmospheric boundary layer, sounds paradoxical in view of the separation between the scales of elementary dunes and giant dunes. Here we perform a wavelet analysis of field-measured giant dune profiles (Fig. 4a, b), which allows a determination of the length scales of the structures that compose these dunes, taking into account the whole relief and not only the horizontal succession of slip-faces. This analysis shows the large-scale modulation of the relief at the wavelength λ_g as well as the covering of elementary dunes at λ_s. It also reveals the presence of superimposed structures at all scales in between. This confirms that superimposed dunes grow progressively by merging, which is possible because the propagation velocity of superimposed bed-forms decreases with their size[10]. Elementary dunes are continuously regenerated because any sufficiently long flat sand area destabilizes linearly at the scale λ_s (ref. 10). Amalgamation results in a global pattern coarsening that follows the amplitude-wavelength curve of Fig. 4c. As the wavelength becomes comparable to the average atmospheric boundary layer depth, dune height is maximized, and the corresponding structures

dominate the atmospheric dynamics as well as the emergence of the pattern. The nonlinear selection of the maximum of the curve $a(\lambda)$ is a generic result of systems whose pattern coarsening results from a screening interaction between neighbouring objects[24] or from a phase-negative diffusion instability[25].

So far, the dynamical mechanisms involved—turbulence, resonant surface waves and dune interaction—are valid for subaqueous dunes[26,27] as well as aeolian ones. In the aeolian case, however, an additional stabilizing mechanism becomes important at small k that is not present in rivers (Fig. 3). It is related to the possibility of radiating internal waves into the atmosphere when the exciting frequency Uk is lower than N. In the subaqueous case this mechanism is not relevant, so that dunes with wavelengths much greater than the flow depth may grow. And indeed, very long river dunes, with wavelengths up to several tens of times the water depth, have been observed in large rivers[26]. We expect their formation mechanism to be of the same nature as for aeolian giant dunes. As on Earth, the atmospheric boundary layers on Mars and Titan[28,29] have kilometre-scale depths. However, the elementary wavelength on Mars is around 600 m (ref. 11), whereas on Titan it is expected to be of the order of a few metres; so although aeolian structures on Mars and Titan are comparable in size[30], only Titan's are giant dunes.

In this exploratory model, we have intentionally omitted several aspects needed to reproduce natural giant dunes in detail: three-dimensional shapes, the Coriolis effect, daily and seasonal atmospheric cycles, erratic winds. The good agreement with wavelength measurements indicates that the stabilizing influence of the free atmosphere is essential to our understanding of the emergence of all giant dune types.

1. Fryberger, S. G. & Dean, G. in *A Study of Global Sand Seas* (ed. McKee, E. D.) 137–169 (Geological Survey Professional Paper 1052, 1979).

2. Werner, B. T. Eolian dunes; computer simulations and attractor interpretation. *Geology* 23, 1107–1110 (1995).

3. Tsoar, H. Dynamic processes acting on a longitudinal (seif) sand dune. *Sedimentology* 30, 567–578 (1983).

4. Bristow, C. S., Bailey, S. D. & Lancaster, N. The sedimentary structure of linear sand dunes. *Nature* 406, 56–59 (2000).

5. Lancaster, N. The dynamics of star dunes: an example from Gran Desierto, Mexico. *Sedimentology* 36, 273–289 (1989).

6. Andreotti, B., Claudin, P. & Douady, S. Selection of dune shapes and velocities. Part 1: Dynamics of sand, wind and barchans. *Eur. Phys. J. B* 28, 321–339 (2002).

7. Wiggs, G. F. S., Livingstone, I. & Warren, A. The role of streamline curvature in sand dune dynamics: evidence from field and wind tunnel measurements. *Geomorphology* 17, 29–46 (1996).

8. Lancaster, N., Nickling, W. G., McKenna-Neuman, C. K. & Wyatt, V. E. Sediment flux and airflow on the stoss slope of a barchan dune. *Geomorphology* 17, 55–62 (1996).

9. Andreotti, B., Claudin, P. & Douady, S. Selection of dune shapes and velocities. Part 2. *Eur. Phys. J. B* 28, 341–352 (2002).

10. Elbelrhiti, H., Claudin, P. & Andreotti, B. Field evidence for surface-wave-induced instability of sand dunes. *Nature* 437, 720–723 (2005).

11. Claudin, P. & Andreotti, B. A scaling law for aeolian dunes on Mars, Venus, Earth, and for subaqueous ripples. *Earth Planet. Sci. Lett.* 252, 30–44 (2006).

12. Kroy, K., Sauermann, G. & Herrmann, H. Minimal model for sand dunes. *Phys. Rev. Lett.* 88, 054301 (2002).

13. Hersen, P., Douady, S. & Andreotti, B. Relevant length scale of barchan dunes. *Phys. Rev. Lett.* 89, 264301 (2002).

14. Bristow, C. S., Duller, G. A. T. & Lancaster, N. Age and dynamics of linear dunes in the Namib desert. *Geology* 35, 555–558 (2007).

15. Kocurek, G., Havholm, K. G., Deynoux, M. & Blakey, R. C. Amalgamated accumulations resulting from climatic and eustatic changes, Akchar Erg, Mauritania. *Sedimentology* 38, 751–772 (1991).

16. Wilson, I. G. Aeolian bedforms—their development and origins. *Sedimentology* 19, 173–210 (1972).

17. Hanna, S. R. The formation of longitudinal sand dunes by large helical eddies in the atmosphere. *J. Appl. Meteorol.* 8, 874–883 (1969).

18. Stull, R. B. *An Introduction to Boundary Layer Meteorology* Ch. 1, 11 (Kluwer Academic, 1988).

19. Lancaster, N. The development of large aeolian bedforms. *Sedim. Geol.* 55, 69–89 (1988).

20. Ewing, R. C., Kocurek, G. & Lake, L. W. Pattern analysis of dune-field parameters. *Earth Surf. Processes Landforms* 31, 1176–1191 (2006).

21. Wurtele, M. G., Sharman, R. D. & Datta, A. Atmospheric lee waves. *Annu. Rev. Fluid Mech.* 28, 429–476 (1996).

22. Baddock, M. C., Livingstone, I. & Wiggs, G. F. S. The geomorphological significance of airflow patterns in transverse dune interdunes. *Geomorphology* 87, 322–336 (2007).

23. Parteli, E. J. R., Schwämmle, V., Herrmann, H. J., Monteiro, L. H. U. & Maia, L. P. Profile measurement and simulation of a transverse dune field in the Lençóis Maranhenses. *Geomorphology* 81, 29–42 (2006).

24. Andersen, K. H., Chabanol, M.-L. & van Hecke, M. Dynamical models for sand ripples beneath surface waves. *Phys. Rev. E* 63, 066308 (2001).

25. Politi, P. & Misbah, C. When does coarsening occur in the dynamics of one-dimensional fronts? *Phys. Rev. Lett.* 92, 090601 (2004).

26. Best, J. The fluid dynamics of river dunes: a review and some future directions. *J. Geophys. Res.* 100, F04S02 (2005).

27. Raudkivi, A. J. Transition from ripples to dunes. *J. Hydraul. Eng.* 132, 1316–1320 (2006).

28. Tokano, T., Ferri, F., Colombatti, G., Mäkinen, T. & Fulchignoni, M. Titan's planetary boundary layer structure at the Huygens landing site. *J. Geophys. Res.* 111, E08007 (2006).

29. Savijärvi, H., Määttänen, A., Kauhanen, J. & Harri, A. M. Mars Pathfinder: New data and new model simulations. *Q. J. R. Meteorol. Soc.* 130, 669–683 (2004).

30. Lorenz, R. D. *et al.* The sand seas of Titan: Cassini RADAR observations of longitudinal dunes. *Science* 312, 724–727 (2006).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.C. (claudin@pmmh.espci.fr).

LETTERS

Devonian arthrodire embryos and the origin of internal fertilization in vertebrates

John A. Long[1,2,3], Kate Trinajstic[4] & Zerina Johanson[5]

Evidence of reproductive biology is extremely rare in the fossil record. Recently the first known embryos were discovered within the Placodermi[1], an extinct class of armoured fish, indicating a viviparous mode of reproduction in a vertebrate group outside the crown-group Gnathostomata (Chondrichthyes and Osteichthyes). These embryos were found in ptyctodontids, a small group of placoderms phylogenetically basal to the largest group, the Arthrodira[2,3]. Here we report the discovery of embryos in the Arthrodira inside specimens of *Incisoscutum ritchiei* from the Upper Devonian Gogo Formation of Western Australia[4] (approximately 380 million years ago), providing the first evidence, to our knowledge, for reproduction using internal fertilization in this diverse group. We show that *Incisoscutum* and some phyllolepid arthrodires possessed pelvic girdles with long basipterygia that articulated distally with an additional cartilaginous element or series, as in chondrichthyans, indicating that the pelvic fin was used in copulation. As homology between similar pelvic girdle skeletal structures in ptyctodontids, arthrodires and chondrichthyans is difficult to reconcile in the light of current phylogenies of lower gnathostomes[2,3,5], we explain these similarities as being most likely due to convergence (homoplasy). These new finds confirm that reproduction by internal fertilization and viviparity was much more widespread in the earliest gnathostomes than had been previously appreciated.

The first appearance of viviparity was considered to have evolved within the chondrichthyans, where today it is the dominant form of reproduction, occurring far less commonly among the osteichthyans[6,7]. However, discovery of viviparity within ptyctodontids, a group belonging in the extinct class Placodermi[1], shifted the evolutionary origin of this reproductive mode back to the most phylogenetically basal gnathostome (jawed vertebrate) group[2,3]. All extant chondrichthyan taxa are sexually dimorphic, with males possessing distinct erectile claspers that are elongate modifications of the pelvic fins used for sperm transfer. Approximately 55% of all chondrichthyans are viviparous[7] and molecular phylogenies indicate that this mode of reproduction arose many times within both chondrichthyans[7,8] and actinopterygians[7,9,10]. Once viviparity develops, the complex physiological requirements that accompany live-bearing generally prohibit reversal back to external spawning[7], although there are rare cases of reversal from live-bearing to egg-laying in squamate reptiles[11]. The arthrodires are the most diverse group of placoderms (>200 species) and they show no signs of sexual dimorphism. Bone morphology and ossification patterns in ptyctodont embryos from *Materpiscis* and *Austroptyctodus*, preserved in three-dimensional uncrushed form from the Upper Devonian Gogo Formation of Western Australia[1] (see Supplementary Information), now allow us to reinterpret small arthrodires discovered within the body cavities of two adult arthrodires from Gogo, *Incisoscutum ritchiei*, as representing embryos (Fig. 1). The small arthrodires were previously presumed to have been

eaten by the larger adult fish[4]. However, the excellent preservation of the small arthrodires does not match poorly preserved stomach contents observed in other placoderms[12–14]. The small *Incisoscutum* specimens were described as 'gut contents' on the basis of the disorganized arrangement of one specimen (P50934); their posterior position behind the trunk shield; presumed gastric etching of some of the trunk plates; and one embryo (P57640) facing posteriorly in relation to the adult[4]. However, the overall position of these embryos (Fig. 1a, b) is exactly the same as the position of embryo 2 within the ptyctodont *Austroptyctodus*[1] (WAM 86.9.662, see Supplementary Information). The disorganized arrangement of dermal plates in P50934 is a result of scattering following the opening of the abdominal cavity. Embryo 3 within *Austroptyctodus* shows a similar disorganized pattern of dermal plates, with some plates outside the abdominal cavity (see Supplementary Information). Worm casts scattered throughout the body and surrounding matrix indicate that the *Incisoscutum* carcass was open to the environment for some time after death (Fig. 1c, d). The posterior orientation of the small arthrodire inside P57640 was considered to be evidence that it was a prey item swallowed head first[4], but this is consistent with the orientation of all four known embryonic ptyctodonts[1]. The stomach contents of *Coccosteus cuspidatus* (Middle Devonian period, Scotland) were described as being an ovoid phosphatic mass, with prey items including the scales of acanthodians and possible dipnoan scales[12]. In addition, detrital grains interpreted as gizzard stones occur inside *Coccosteus cuspidatus*[14]. The complete preservation of the dermal plates and the absence of other taxa associated with the embryonic plates suggest that the arthrodires within adult *I. ritchiei* specimens can now be excluded as stomach contents.

The presumed gastric 'acid etching' on the dermal plates of the arthrodire embryos, forming an 'open weave' appearance on the bone surface[4], can now be attributed to an early stage of ossification. Differential ossification is also seen in the juveniles of ichthyosaurs[15], pterosaurs[16], the placoderms *Bothriolepis*[17] and *Asterolepis*[18], and the ptyctodont embryos[1]. Juveniles (13–14 mm) of the antiarch *Asterolepis ornata* show lamellar bone with an irregular network of bony trabeculae on the external surface of the lamina. The second stage of development is the formation of the spongious middle layer[18]. Larger individuals (longer than 30 mm) have fine meshed reticular ornament, which sometimes bears tubercles. The ptyctodont and *Incisoscutum* embryos appear to have reached this final developmental stage, although, unlike *Asterolepis*, they remained *in utero*.

Further new evidence of arthrodire reproductive biology has come to light with our re-examination of the pelvic girdle structure in *Incisoscutum* and in the Phyllolepida, a dorsoventrally flattened group of basal arthrodires (Fig. 2). The pelvic girdle is well known in *Incisoscutum* and the perichondrally ossified pelvic radialia are well preserved[4]. The posterior large facet for the 'metapterygium' identified previously[4] is here reinterpreted as being for a long, posteriorly
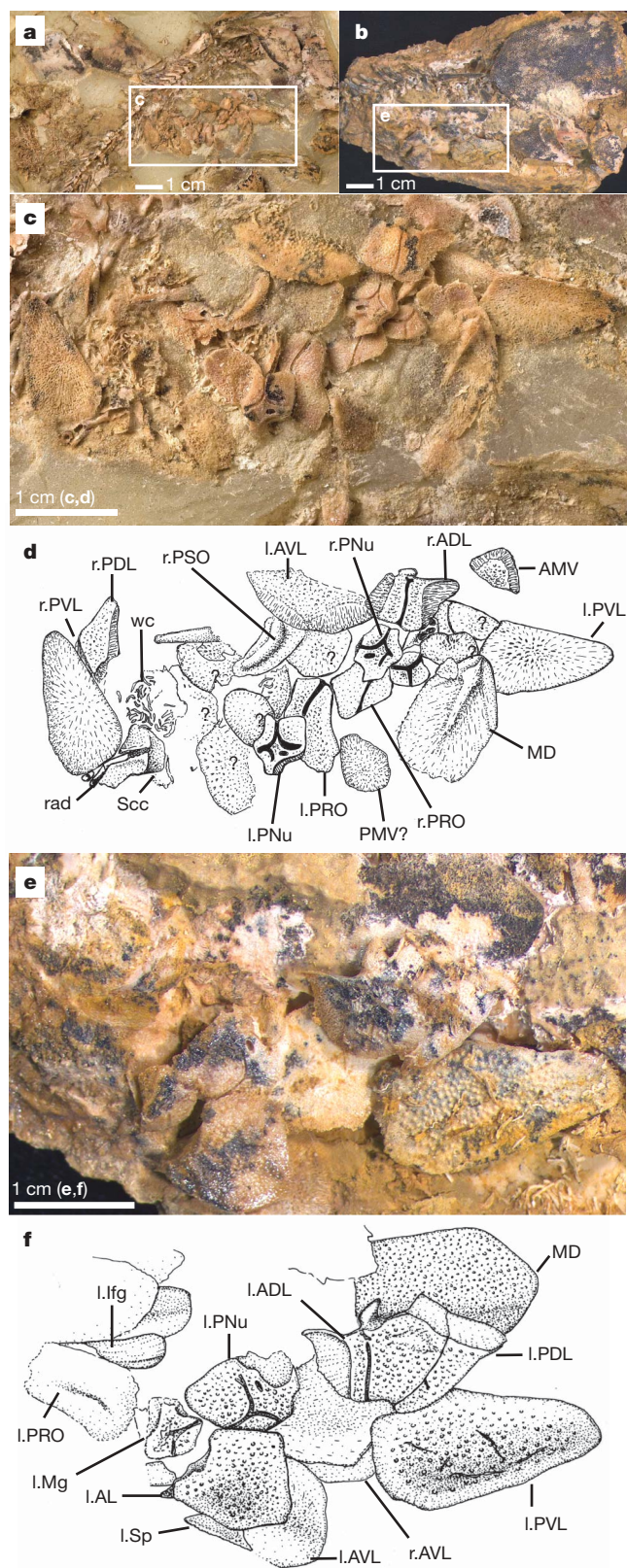
Figure 1 | *Incisoscutum ritchiei* specimens showing embryo fossils.
a, c, d, P50934; **b, e, f**, P57640. AMV, anterior median ventral; l.AL, left
anterolateral; l.AVL/r.AVL, left and right anterior ventrolaterals; l.Ifg, left
inferognathal; l.Mg, left marginal; l.PDL, left posterior dorsolateral; l.PVL/
r.PVL, left and right posterior ventrolaterals; l.Sp, left spinal; MD, median
dorsal; PMV?, posterior median ventral?; r.ADL, right anterior dorsolateral;
l.PNu/r.PNu, left and right paranuchals; r.PDL, right posterior dorsolateral;
l.PRO/r.PRO, left and right preorbitals; r.PSO, right postsuborbital; rad,
radialia; Scc, scapulocoracoid; wc, worm casts.
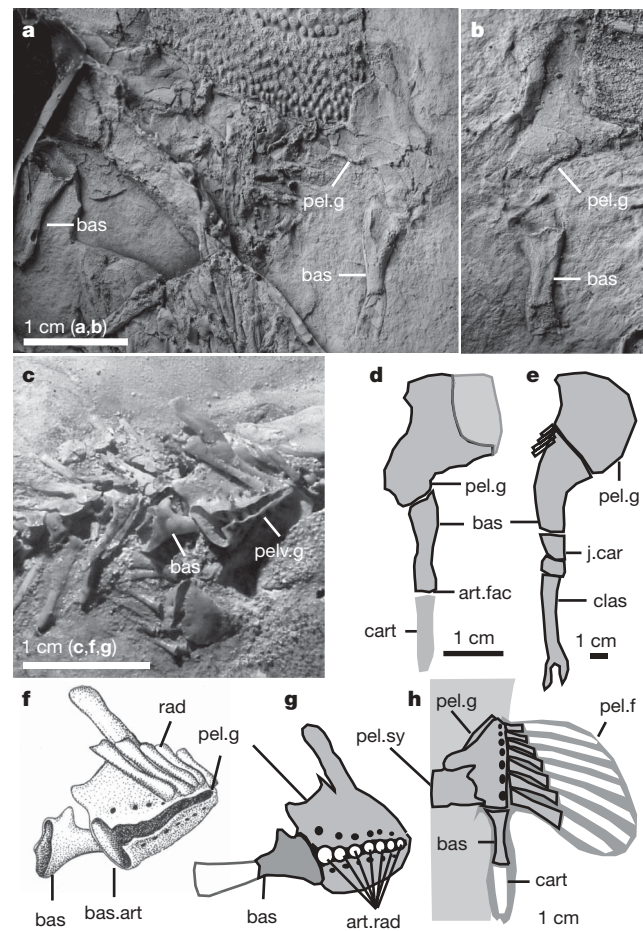
Figure 2 | Placoderm and shark pelvic structures. a, **b**, **d**, The phyllolepid
Austrophyllolepis ritchiei (MV P160746) showing pelvic girdle in ventral
(**a**) and dorsal (**b**) aspects, and compared, after being restored (**d**), with the
shark *Cobelodus* (**e**). **c**, **f**–**h**, *Incisoscutum ritchiei*. **c**, **f**, **g**, WAM 04.10.02.
g, Pelvic girdle restored (full extent of basipterygium unknown). **h**, Pelvic
fin structures restored in dorsal view showing restored fin cartilages in white.
art.fac, articular facet for additional pelvic elements; art.rad, radial
articulations; bas, basipterygium; bas.art, basipterygial articulation facet;
cart, cartilage distal to basipterygium; clas, clasper; j,car, junction cartilages;
pel.f, pelvic fin; pel.g, pelvic girdle; pel.sy, pelvic symphysis; rad, radialia.

facing basipterygium (Fig. 2c, f) by comparison with the pelvic girdle
in selachians (Fig. 2e). The basipterygium is seen in WAM 04.10.2
(Fig. 2c, f) and is here reconstructed as in life (Fig. 2g, h).

The pelvic girdle in phyllolepids was first described from
Austrophyllolepis from the Middle Devonian strata of Victoria,
Australia[19]. It was compared with claspers in the stem chondrich-
thyan *Cobelodus*, but dismissed owing to lack of demonstrable sexual
dimorphism. We now identify a broad-based pelvic plate articulating
to a long, posteriorly directed basipterygium in these phyllolepids
(Fig. 2a, b, d). The distal end of the basipterygium has an articulation
surface for an additional cartilaginous segment or series. Further
complete pelvic girdles of similar gross morphology occur in another
phyllolepid, *Cowralepis*, from New South Wales, Australia[20]. The
long basipterygia of *Cowralepis*, *Austrophyllolepis* and *Incisoscutum*
all show a remarkably chondrichthyan-like morphology (Fig. 2e). A
comparable structure is unknown in early osteichthyan pelvic girdles,
although juvenile sturgeons do show a series of pelvic basals in similar
topological position to the basipterygium linking the last radial with
the metapterygium[21]. In view of the presence of embryos within
Incisoscutum, in association with a long basipterygium, this structure
probably had a role in copulation as it does in chondrichthyans. It
must have articulated distally with at least one other unossified car-
tilage, possibly for a clasper in males and/or a non-clasping cartilage

in females. The proximal part of the pelvic girdle (pelvic basal plate and basipterygium) appears to be the same shape and size in both male and female phyllolepids, comparable to modern chondrichthyan pelvic girdles. In *Squalus* the pectoral girdle and basipterygium are of the same shape and size in both sexes, so only the distal elements differ in morphology. This loss of ossification in the distal pelvic fin skeleton most probably corroborates with the use of the pelvic fin in erectile copulation, as occurs in chondrichthyans.

Viviparity has been reported in several fossil fish groups including the Middle Triassic actinopterygian *Peltopleurus lissocephalus* Brough 1939[22] the coelacanth *Holophagus* Watson 1927[23] and the holocephalan *Delphyodontos dacriformes* Lund 1980[24]. This indicates that viviparity developed within Osteichthyes without claspers, and often with no signs of external sexual dimorphism. The presence of embryos in the arthrodire *Incisoscutum* indicate that internal fertilization was the main reproductive mode for this species, and that viviparity could have evolved in other eubrachythoracid arthrodires.

Ptyctodontid claspers are dermal bones[25] having a groove for cartilaginous internal support that has been suggested as being homologous with the cartilaginous basipterygium of elasmobranch claspers[26]. We have no evidence, however, of sexual dimorphism shown in the pelvic fins of arthrodires, but this is not a requirement for internal fertilization, as shown by the various families of extant viviparous teleosteans and the sarcopterygian *Latimeria*[23].

Mapping of viviparity and the presence of claspers onto recent hypotheses of gnathostome interrelationships shows that there is no parsimonious reconciliation of these features with current phylogenetic schemes (Fig. 3a, b). Claspers in ptyctodontids and chondrichthyans are thus here interpreted as being non-homologous, as to accept homology based on these phylogenies[2,3] implies either a loss of claspers in osteichthyans (and/or some acanthodians[3]), or a monophyletic grouping of placoderms and chondrichthyans, which assumes claspers might have been present in other placoderm groups[27]. We leave the problem unresolved, and accept that the most likely explanation for this similarity in pelvic fin morphology in ptyctodontids and chondrichthyans is homoplasy, or convergence.

Our discovery of the first embryos in arthrodiran fishes thus demonstrates that internal fertilization was a more widespread reproductive strategy in the earliest jawed vertebrates. Furthermore, our observation that certain arthrodires (including phyllolepids) had elongated pelvic basipterygia that articulated distally with additional cartilage elements as in chondrichthyans supports the view that a copulatory mode of reproduction involving manipulation of the pelvic fin for sperm delivery had evolved at this early stage of vertebrate evolution. Although homology between such structures in placoderms (ptyctodontids) and chondrichthyans is not yet resolved, such observations provide informative new data to contribute to the ongoing debate concerning the origins and interrelationships of the first jawed vertebrates.

METHODS SUMMARY

The specimens of *Incisoscutum* were prepared at the Natural History Museum, London, using the standard acetic acid technique. Reference 1 provides details of this method as applied to preparing out embryonic ptyctodontid placoderms. The specimens of *Austrophyllolepis* were prepared using negative preparation by immersing the specimens in 10% HCl overnight to dissolve away the weathered bone. They were then washed in water for 24 h, dried and hardened using 3% paraloid in acetone. A latex peel (darkened with Indian ink) of the cleaned mould was then made, and whitened with ammonium chloride sublimate for high-contrast photography.

1. Long, J. A., Trinajstic, K., Young, G. C. & Senden, T. Live birth in the Devonian period. *Nature* **453**, 650–652 (2008).
2. Goujet, D. & Young, G. C. in *Recent Advances in the Origin and Early Radiation of Vertebrates* (eds Arratia, G., Wilson, M. V. H. & Cloutier, R.) 109–126 (Dr Freiderich Pfeil, 2004).
3. Brazeau, M. D. The braincase and jaws of a Devonian 'acanthodian' and modern gnathostome origins. *Nature* **457**, 305–308 (2009).
4. Dennis, K. & Miles, R. S. A pachyosteomorph arthrodire from Gogo, Western Australia. *Zool. J. Linn. Soc.* **73**, 213–258 (1981).
5. Janvier, P. in *Major Events in Early Vertebrate Evolution* (ed. Ahlberg, P. E.) 172–186 (Taylor & Francis, 2001).
6. Wourms, J. Viviparity: The maternal-fetal relationship in fishes. *Am. Zool.* **21**, 473–515 (1981).
7. Blackburn, D. in *Viviparous Fishes* (eds Grier, H. & Uribe, M.) 287–301 (New Life Publications, 2005).
8. López, J. A., Ryburn, J. A., Fedrigo, O. & Naylor, G. J. P. Phylogeny of sharks of the family Triakidae (Carcharhiniformes) and its implications for the evolution of carcharhiniform placental viviparity. *Mol. Phyl. Evol.* **40**, 50–60 (2006).
9. Mank, J. E. & Avise, J. C. Supertree analyses of the roles of viviparity and habitat in the atherinomorph fishes. *Eur. Soc. Evol. Biol* **19**, 734–740 (2006).
10. Webb, S. *et al.* Molecular phylogeny of the livebearing Goodeidae (Cyprinidontiformes). *Mol. Phyl. Evol.* **30**, 527–544 (2004).
11. Lee, M. Y. S. & Shine, R. Reanalysis of the evolution of viviparity and egg-guarding in squamate reptiles. *Herpetologica* **55**, 538–549 (1998).
12. Davidson, R. & Trewin, N. Unusual preservation of the internal organs of acanthodian and actinopterygian fish in the Middle Devonian of Scotland. *Scot. J. Geol.* **41**, 129–134 (2005).
13. Long, J. A. *Swimming in Stone—The Amazing Gogo Fossils of the Kimberley* 320 (Fremantle Arts Centre Press, 2006).
14. Trewin, N. H. Palaeoecology and sedimentology of the Achanarras fish bed of the Middle Old Red Sandstone, Scotland. *Trans. R. Soc. Edinb. Earth Sci.* **77**, 21–46 (1986).
15. Maxwell, E. E. & Caldwell, M. W. First record of live birth in Cretaceous ichthyosaurs: closing an 80 million year gap. *Proc. R. Soc. Lond. B* **270** (suppl.), S104–S107 (2003).
16. Chiappe, L. M., Cordorniu, L., Greller-Tinner, G. & Rivarola, D. Argentinian unhatched pterosaur fossil. *Nature* **432**, 571–572 (2004).
17. Young, G. C. Antiarchs (placoderm fishes) from the Devonian Aztec Siltstone, southern Victoria Land, Antarctica. *Palaeontographica* A202, 1–125 (1988).
18. Upeniece, I. The unique fossil assemblage from the Lode Quarry (Upper Devonian, Latvia). *Mitt. Mus. Natkd. Berl. Geowiss* **4**, 101–119 (2001).
19. Long, J. A. New phyllolepid placoderms from the Late Devonian of Victoria, Australia and the relationships of the group. *Proc. Linn. Soc. NSW* **107**, 263–308 (1984).
20. Ritchie, A. *Cowralepis*, a new genus of phyllolepid fish (Pisces, Placodermi) from the Late Middle Devonian of New South Wales, Australia. *Proc. Linn. Soc. NSW* **126**, 215–259 (2005).
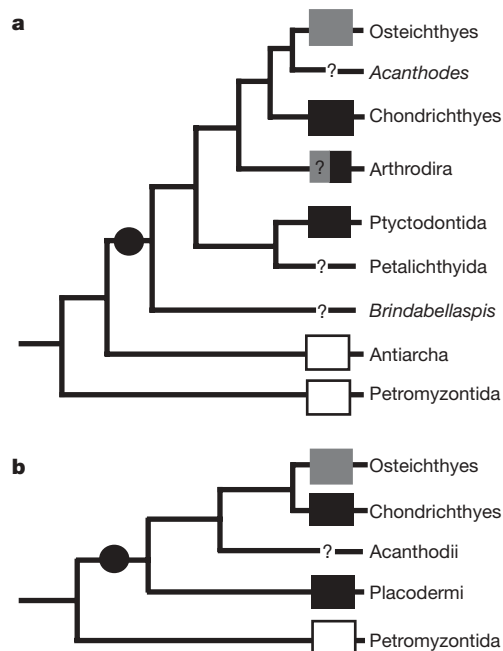

Figure 3 | Lower gnathostome phylogenies showing reproductive strategies. Viviparity has arisen many times in vertebrate evolution using different morphologies for internal fertilization. **a**, After Brazeau[3]. **b**, After Young[26] and Goujet and Young[2] (based on two broad assumptions: (1), placoderm monophyly; (2), all placoderms reproduced by internal fertilization). Black circle, presence of pelvic fins; black rectangle, viviparity with claspers present; grey rectangle, spawning and/or internal fertilization without claspers; white rectangle, spawning mode of reproduction; ?, pelvic region too poorly known to infer reproductive mode.

21. Sewertzoff, A. N. Development of the pelvic fins of *Acipenser ruthenus*. New data for the theory of the paired fins of fishes. *J. Morph. Physiol.* **41**, 547–579 (1926).
22. Bürgin, T. Reproduction in the Middle Triassic actinopterygians; complex fin structures and evidence of viviparity in fossil fishes. *Zool. J. Linn. Soc.* **100**, 379–391 (1990).
23. Schultze, H.-P. Early growth stages in coelacanth fishes. *Nat. New Biol.* **236**, 90–91 (1972).
24. Lund, R. Viviparity and interuterine feeding in a new holocephalan fish from the Lower Carboniferous of Montana. *Science* **209**, 697–699 (1980).
25. Miles, R. S. Observations on the ptyctodont fish, *Rhamphodopsis* Watson. *Zool. J. Linn. Soc.* **47**, 99–120 (1967).
26. Young, G. C. The relationships of placoderm fishes. *Zool. J. Linn. Soc.* **88**, 1–57 (1986).
27. Stensiö, E. A. Anatomical Studies on the Arthrodiran Head. *Kungl. Svenska Vetensk. Akad. Handl.* 4 (9) 2, 1–419 (1963).

LETTERS

Cellular prion protein mediates impairment of synaptic plasticity by amyloid-β oligomers

Juha Laurén[1], David A. Gimbel[1], Haakon B. Nygaard[1], John W. Gilbert[1] & Stephen M. Strittmatter[1]

A pathological hallmark of Alzheimer's disease is an accumulation of insoluble plaque containing the amyloid-β peptide of 40–42 amino acid residues[1]. Prefibrillar, soluble oligomers of amyloid-β have been recognized to be early and key intermediates in Alzheimer's-disease-related synaptic dysfunction[2–9]. At nanomolar concentrations, soluble amyloid-β oligomers block hippocampal long-term potentiation[7], cause dendritic spine retraction from pyramidal cells[5,8] and impair rodent spatial memory[2]. Soluble amyloid-β oligomers have been prepared from chemical syntheses, transfected cell culture supernatants, transgenic mouse brain and human Alzheimer's disease brain[2,4,7,9]. Together, these data imply a high-affinity cell-surface receptor for soluble amyloid-β oligomers on neurons—one that is central to the pathophysiological process in Alzheimer's disease. Here we identify the cellular prion protein (PrP^C) as an amyloid-β-oligomer receptor by expression cloning. Amyloid-β oligomers bind with nanomolar affinity to PrP^C, but the interaction does not require the infectious PrP^{Sc} conformation. Synaptic responsiveness in hippocampal slices from young adult PrP null mice is normal, but the amyloid-β oligomer blockade of long-term potentiation is absent. Anti-PrP antibodies prevent amyloid-β-oligomer binding to PrP^C and rescue synaptic plasticity in hippocampal slices from oligomeric amyloid-β. Thus, PrP^C is a mediator of amyloid-β-oligomer-induced synaptic dysfunction, and PrP^C-specific pharmaceuticals may have therapeutic potential for Alzheimer's disease.

To characterize amyloid-β-oligomer-binding sites, we synthesized amyloid-β(1–42) peptide conjugated to biotin through the α amino group (biotin–Aβ42), and then denatured the peptide and allowed oligomers to form as described for amyloid-β-derived diffusible ligands[4]. Consistent with findings for untagged Aβ42 oligomers[5], biotin–Aβ42-oligomer preparations contain spherical particles of 5–6 nm in diameter visible by negative staining with transmission electron microscopy, with rare protofibrils and no larger fibrils (Fig. 1a). Approximately 50% of the peptide migrates by size-exclusion chromatography (SEC) as a distinct assembly with a size of approximately 500 kDa corresponding to 50–100 amyloid-β monomers (Fig. 1b). Low-molecular-weight forms of Aβ42 in either oligomeric or fresh preparations migrate by SEC as monomers (Fig. 1b), demonstrating that the trimers or tetramers observed by SDS–polyacrylamide gel electrophoresis (Supplementary Fig. 1) are not present under native conditions[10]. Aβ42 oligomer binds to hippocampal neurons, whereas freshly prepared biotin–Aβ42 does not (Fig. 1c and Supplementary Fig. 2). Biotin–Aβ42-oligomer binding is enriched in MAP2 (also known as MTAP2)-positive dendrites, with lower levels in βIII-tubulin (TUBB3)-positive axons, and very low levels in astroglial cells (Supplementary Fig. 3a, c, data not shown and ref. 6). The Aβ42-oligomer binding is most concentrated at post-synaptic densities marked by immunoreactive PSD-95 (also known as DLG4; Supplementary Fig. 3b). Binding to neurons is saturable, with an apparent dissociation constant (K_d) of 50–100 nM monomer equivalent

(Fig. 1d). The K_d of the relevant Aβ42 assembly must be much less than 100 nM because minimal binding is detected with freshly prepared Aβ42. If the Aβ42 species responsible for binding contains 100 monomers and represents 50% of all biotin–Aβ42 in the preparation, the corrected affinity would be ~0.4 nM. Although this formulation of Aβ42 oligomer is not chromatographically identical to the Aβ42 oligomer from brain[2,3,9], it affords detection of high-affinity binding sites likely to share pathological actions with sites for other Aβ42-oligomer preparations[5,6,11].

A key requirement for expression cloning of Aβ42-oligomer-binding sites is the existence of a cell line with low background binding. COS-7 cells exhibit <5% of the biotin–Aβ42-oligomer binding level in hippocampal neurons. We expressed complementary DNAs from an adult mouse brain library in COS-7 cells and screened for biotin–Aβ42-oligomer binding. From 225,000 clones, 2 independent positive clones were isolated and both were found to encode full-length mouse PrP (Fig. 1e). Aβ42 oligomers bind to cells expressing the PrP^C conformation; interaction is not dependent on the PrP^{Sc} conformation required for infectious prion disease[12]. PrP^C is known to interact with copper ion but this does not alter Aβ42-oligomer binding (Supplementary Fig. 4). Like hippocampal neurons, PrP^C-expressing COS-7 cells have much lower affinity for freshly prepared low-molecular-weight biotin–Aβ42 (Fig. 1e and Table 1). The apparent dissociation constant for biotin–Aβ42-oligomer binding to PrP^C-expressing COS-7 cells is indistinguishable from that for biotin–Aβ42-oligomer binding to hippocampal neurons (Fig. 1f, g and Table 1). The selectivity of PrP for binding Aβ42 oligomer versus fresh Aβ42 is reflected in the ratio of K_d values and must be greater than 20 (>2,000 nM/92 nM) based on the total peptide monomer concentration in the Aβ42-oligomer preparation (Table 1), or as great as 5,000 (>2,000 nM/0.4 nM) based on the molar concentration of Aβ42 oligomer estimated by SEC/light scattering (LS) (Fig. 1b).

To explore any contribution of the biotin tag to PrP affinity, we prepared untagged Aβ42 oligomer and examined binding to PrP^C-expressing cells with an anti-amyloid-β antibody (Supplementary Fig. 5). Untagged Aβ42 binding is localized to PrP^C-expressing cells. Thus, binding is mediated by the amyloid-β amino acid residues. The simplest model for PrP^C expression inducing Aβ42-oligomer binding is a direct interaction between the two polypeptides. To verify this, we examined the interaction of purified PrP–Fc with Aβ42 (Supplementary Fig. 6). A control Fc protein, immobilized on a resin, retained neither freshly prepared nor oligomeric preparations of Aβ42. In contrast, PrP–Fc protein retained Aβ42 peptide through a direct physical interaction. The pre-incubated oligomeric form of Aβ42 was retained to a 2.5-fold greater degree than the freshly prepared peptide. The preference of PrP for amyloid-β oligomer versus fresh Aβ42 is less complete here than in the cellular assays, perhaps owing to the use of concentrated solid-phase purified reagents and higher concentrations.

---

[1]Cellular Neuroscience, Neurodegeneration and Repair Program, Yale University School of Medicine, New Haven, Connecticut 06536, USA.
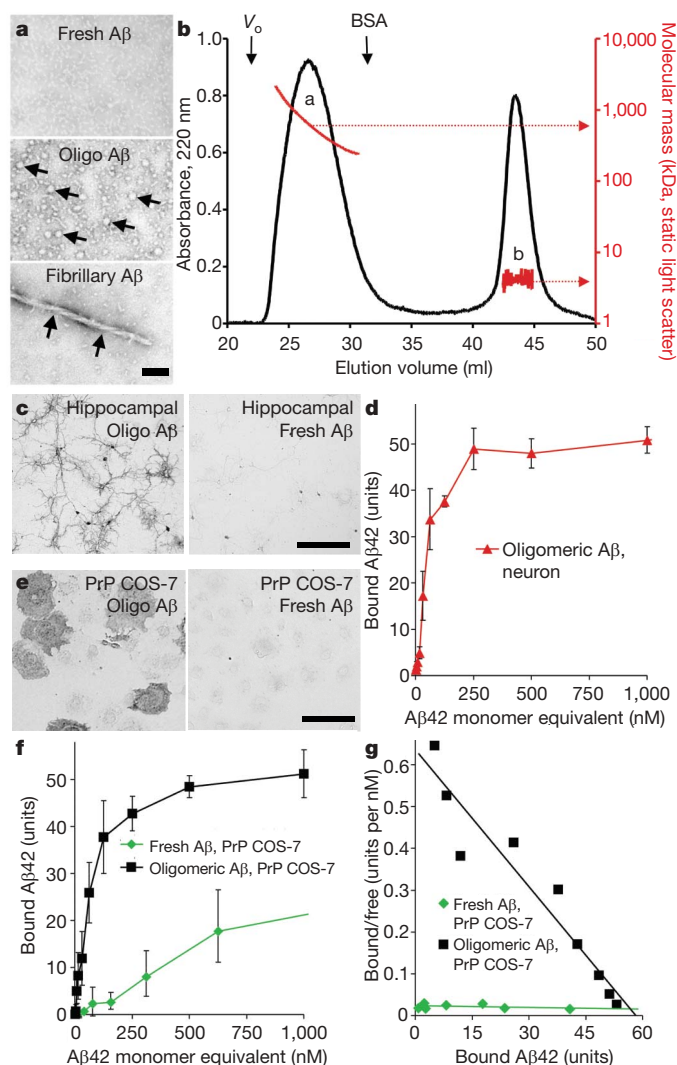
Figure 1 | Oligomeric Aβ42 binds to neurons and to cells expressing PrP^C.
a, Freshly prepared, oligomeric or fibrillary preparations of Aβ42 were examined by transmission electron microscopy with negative staining. The arrows indicate globular oligomers in the middle segment and a fibril in the lower segment. Aβ, amyloid-β. Scale bar, 25 nm. **b**, Oligomeric Aβ42 peptide was analysed by SEC, monitoring absorbance at 220 nm (black) and light scattering (red). The void volume (V_o) and elution of bovine serum albumin (BSA) from a separate run are shown. Peak 'a' is oligomeric Aβ42 and 'b' is monomeric Aβ42. **c**, Oligomeric Aβ42 peptide (200 nM total peptide) binds to 21 DIV hippocampal neurons (left), whereas fresh Aβ42 (200 nM) does not (right). Bound biotin–Aβ42 was visualized by alkaline-phosphatase-conjugated streptavidin. **d**, Dose-dependence of oligomeric Aβ42 binding to hippocampal neurons. **e**, The binding of 40 nM oligomeric or freshly prepared Aβ42 to COS-7 cells expressing PrP^C. **f, g**, Fresh or oligomeric Aβ42 binding to PrP^C-expressing COS-7 cells as a function of Aβ42 total concentration (monomer equivalent for oligomer preparations). Data are mean ± s.e.m., and the Scatchard analysis is presented in **g**. Scale bars, 100 μm for **c** and **e**.

Although the PrP cDNA is the only clone to support oligomeric Aβ42 binding isolated from the brain cDNA library, we considered whether other amyloid-β-binding sites might exist. First, we examined two clones sharing sequence similarity with PrP—doppel (also known as PRND) and SPRN—but neither showed affinity for oligomeric Aβ42 (Table 1 and Supplementary Fig. 7). Second, we screened a pre-existing collection of 352 cDNAs encoding transmembrane proteins one by one. In this format, weaker affinity interactions are detectable than in the initial pooled brain library screen. Amyloid-β precursor-like protein 1 (APLP1) and transmembrane protein 30B (TMEM30B) were isolated through this focused screen and demonstrate K_d values for

oligomeric Aβ42 of 660 nM and 720 nM, respectively (Table 1 and Supplementary Fig. 7). These lower affinity binding proteins exhibit limited specificity for oligomeric Aβ42, as compared to fresh Aβ42. APLP1 shares similarity with amyloid-β precursor protein (APP) and with APLP2, but neither of these proteins binds Aβ42 (Table 1). TMEM30B is similar to TMEM30A, which is expressed at high levels in the brain. TMEM30A supports Aβ42 binding with an affinity similar to TMEM30B and shows no preference for oligomeric species (Table 1). The receptor for advanced glycation end products (RAGE) and the α7 nicotinic acetylcholine receptor (nAChRα7, also known as CHRNA7) have been reported to bind amyloid-β[13,14]. In this heterologous COS-7 cell binding assay, expression of RAGE yielded less Aβ42-oligomer-binding signal than did PrP, APLP1 or TMEM30B, and we failed to detect binding to nAChRα7 (Supplementary Fig. 7). Thus, although several proteins exhibit Aβ42 binding, only PrP has high affinity and high selectivity for the oligomeric peptide.

Binding of Aβ42 oligomer to neurons depends on the developmental stage, with minimal binding to neurons immediately after dissociation from E18 hippocampus. Binding of Aβ42 oligomer to neurons does not become robust until 15–20 days have elapsed *in vitro* (Fig. 2a). The immunoblot level of PrP^C expression closely matches this developmental pattern (Fig. 2b). Immunocytochemically, PrP^C expression is largely restricted to MAP2-positive dendrites of differentiated neurons (Supplementary Fig. 3d and Fig. 2c). Furthermore, localization of PrP immunoreactivity and Aβ42-oligomer binding overlap extensively (Fig. 2c). If PrP^C (encoded by *Prnp*) were the only cellular binding site for Aβ42 oligomers, then no binding would be detected in cultures from *Prnp*^−/− mice at 20 days *in vitro* (DIV). Because we observe 50% reduction of punctate Aβ42-oligomer binding in such cultures (Fig. 2d, e), PrP^C cannot be the only cell-surface molecule binding Aβ42 oligomers. Multiple alternative sites, including APLP1, TMEM30A, TMEM30B, RAGE and unidentified proteins, may explain Aβ42 binding to *Prnp*^−/− neurons.

Different domains of PrP^C have been associated with various activities. The amino-terminal octapeptide repeat domain (amino acids 60–95) contributes to extracellular copper ion binding[15,16]. The unstructured central domain (amino acids 95–134) includes a charge cluster (amino acids 95–110) and a segment with hydrophobic character (amino acids 112–134). This central domain has been implicated in masking a neurodegenerative activity of PrP^C (refs 17 and 18). The carboxy-terminal domain is globular (amino acids 134–231)[19] and the protein is glycosyl phosphatidylinositol (GPI)-anchored to the plasma membrane. We mapped Aβ42-oligomer binding using PrP deletion mutants (Fig. 3). Each mutant protein was expressed at the COS-7 surface by live anti-PrP immunostaining (Fig. 3a and Supplementary Fig. 8). Deletion of the octapeptide repeat domain and the central domain (Δ32–121) abrogates binding, indicating that the globular domain alone cannot mediate binding. The hydrophobic 105–125 region is not a major determinant, because Δ105–125 protein binds Aβ42 oligomers indistinguishably from full-length PrP^C, and because the Δ32–106 variant behaves like the Δ32–121 variant, having no Aβ42-oligomer affinity. To distinguish whether the 95–110 charge cluster or the octapeptide repeat domain is crucial for Aβ42 binding, a mutant lacking the 52–91 segment was expressed. The Δ52–91 mutant exhibits significant Aβ42 binding, implicating the 95–110 region as a principal site for Aβ42-oligomer binding. Consistent with this hypothesis, deletion of 11 amino acids in the Δ95–105 variant reduced binding by 80%, and there was no further reduction in the Δ70–105 variant.

As an alternative method to localize Aβ42 binding within PrP^C, we used anti-PrP antibodies (Fig. 3b–d). Of the six antibodies initially tested, only one (6D11) blocked the binding of Aβ42 assemblies to PrP^C, and did so with a half-maximal inhibitory concentration of 1 nM (Fig. 3b–d and Supplementary Figs 8–10). The 6D11 blockade is epitope-specific because the 7D9 antibody binds avidly to a different epitope but fails to block Aβ42 binding (Fig. 3a–d and Supplementary Figs 8 and 9). The epitope for 6D11 corresponds to amino acids 93–109 of mouse PrP^C, matching the conclusion that the 95–105 amino

Table 1 | Aβ42 binding to COS-7 cells expressing the indicated proteins

Surface protein	Clone source	Oligomer amyloid-β K_d (nM)*	Fresh amyloid-β K_d (nM)*	Oligomer selectivity†
PrP	225,000 brain cDNAs	92 ± 14	>2,000‡	>21
SPRN	Similar to PrP	>2,000‡	>2,000‡	
doppel	Similar to PrP	>2,000	>2,000	
APLP1	352 transmembrane	660 ± 30	>2,000‡	>3
APLP2	Similar to APLP1	>2,000	>2,000	
APP	Similar to APLP1	>2,000	>2,000	
TMEM30A	Similar to TMEM30B	990 ± 60	1,100 ± 60	1.1
TMEM30B	352 transmembrane	720 ± 20	870 ± 70	1.2

APLP1 and TMEM30B were identified by examination of individual clones from a 352-member pre-existing collection of expression vectors for transmembrane proteins (Origene).
* K_d values are mean ± s.e.m. and are based on monomer equivalent concentrations.
† Oligomer selectivity is the ratio of the monomer K_d to the oligomer K_d.
‡ Binding of Aβ42 was detected at 2 μM ligand, but binding was not saturated.
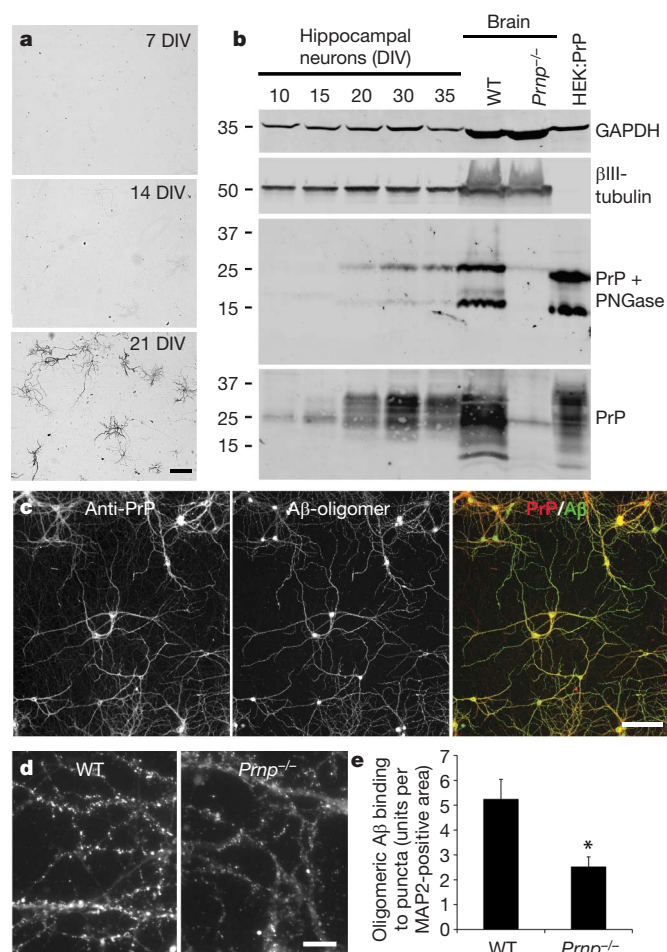
Figure 2 | Characterization of Aβ42 oligomer binding sites. a, Oligomeric Aβ42 (30 nM) binding to hippocampal neurons after culture for the indicated time. Neuronal cell density is similar in the three panels. Scale bar, 100 μm. **b,** Total protein (20 μg) from hippocampal cultures, from whole brain of the indicated genotype or from HEK293T cells transfected with a PrP expression vector was analysed by immunoblot with anti-PrP antibody (8H4), with anti-βIII-tubulin antibody or with anti-GAPDH antibody. Samples for the middle panel were pretreated with endoglycosidase (PNGase F) before gel electrophoresis through a 4–20% polyacrylamide gel in Tris-glycine-SDS. Molecular weight standards are shown at the left in kDa. **c,** Hippocampal neurons from E18 mice after 21 DIV were incubated with biotinylated Aβ42 oligomer (130 nM monomer equivalent) for 1 h at 37 °C and then fixed. Bound amyloid-β was detected with fluorescent avidin (green), and PrP^C with anti-PrP immunocytochemistry (red). Scale bar, 100 μm. **d,** Cultures were prepared from wild-type (WT) or Prnp$^{-/-}$ mice, and then binding of Aβ42-oligomer (130 nM monomer equivalent) was detected as in **c**. Scale bar, 10 μm. **e,** Binding of 100 nM Aβ42 oligomers to puncta in hippocampal cultures as shown in **d**. Data from $n = 6$ pairs of cultures from wild-type and Prnp$^{-/-}$ embryos. Mean and s.e.m.; *$P < 0.05$, two-tailed t-test.

acid region is a primary determinant for binding. To confirm this hypothesis, we examined the effect of an additional antibody (8G8) with an overlapping epitope, amino acids 95–110. The 8G8 antibody blocked the Aβ42–PrP^C interaction, although with a lesser potency than 6D11. The effect of 6D11 was not caused by internalization of PrP^C, because similar cell-surface levels of PrP^C were detectable after 6D11 pre-incubation (Supplementary Fig. 11). The 6D11 antibody is highly specific for PrP^C, as no immunoreactivity was observed in Prnp null brain sections nor was there any reactivity to Aβ42 (Supplementary Fig. 12 and not shown). We conclude that the 95–105 segment of PrP^C contributes to Aβ42-oligomer binding in a 6D11-sensitive manner.

Although these data demonstrate that PrP^C is a high-affinity binding site for Aβ42 oligomers, they do not assess its role in the pathological actions of Aβ42. It has been noted that soluble Aβ42 oligomers suppress long-term potentiation (LTP) of the Schaffer collateral pathway between hippocampal CA3 and CA1 pyramidal cells[7,11]. Therefore, we compared the effects of soluble Aβ42 oligomers on LTP from slices of wild-type versus Prnp$^{-/-}$ mice[20,21]. As reported previously, soluble Aβ42 oligomers (500 nM total peptide, estimated 2 nM Aβ42 oligomer) reduce LTP in hippocampal slices from wild-type mice (Fig. 4a, d). The slope of the excitatory postsynaptic potential (EPSP) after theta burst stimulation is augmented by 80% in control slices, but only by 20% in slices pre-incubated with Aβ42-oligomer preparations. In slices from 2–6-month-old PrP null mice without Aβ42 treatment, Schaffer collateral LTP is indistinguishable from baseline levels of wild-type mice (Fig. 4b), as described previously[22,23]. Notably, there is no inhibition of LTP by Aβ42 oligomers in the Prnp$^{-/-}$ slices (Fig. 4b–d).

The lack of Aβ42 sensitivity for LTP in Prnp$^{-/-}$ slices indicates that PrP^C acts as a receptor for Aβ42 oligomers mediating inhibition of LTP in wild-type slices. Alternatively, chronic loss of PrP^C may lead to developmental and/or compensatory effects that account indirectly for Aβ42 oligomer ineffectiveness. To separate these possibilities, we pretreated wild-type slices with the 6D11 anti-PrP antibody (100 nM for 20 min), which was shown to block Aβ42 binding acutely (Fig. 3). Pretreatment with control immunoglobulin G did not reduce the suppression of LTP by Aβ42 oligomer (Fig. 4d). In contrast, the 6D11-pretreated wild-type slices were protected from LTP suppression by the later addition of Aβ42-oligomer preparations (Fig. 4d). Thus, we conclude that PrP^C exerts a receptor action acutely to mediate Aβ42-oligomer inhibition of synaptic plasticity in the hippocampal slice.

The principal finding of this study is that PrP^C functions as a receptor to mediate the deleterious effects of the Aβ42 oligomer. This hypothesis is supported by our isolation of PrP^C as an Aβ42-oligomer-binding site in an unbiased genome-wide screen, by the match between PrP^C expression and the properties of Aβ42-oligomer-binding sites, and by the localization of amyloid-β binding to a neurodegeneration-associated domain of PrP^C. Although PrP^C is not the sole binding site for amyloid-β oligomers on hippocampal neurons, it is essential for Aβ42-oligomer inhibition of hippocampal LTP. Several publications indirectly support coupling of Aβ42 oligomers and PrP^C. For example,
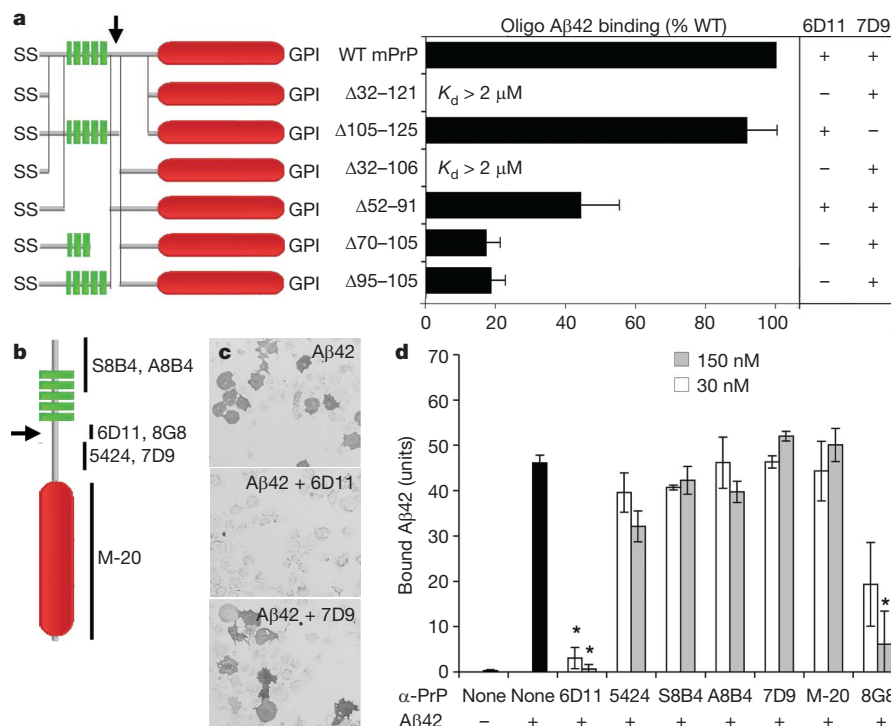
Figure 3 | Aβ42 oligomers bind to residues 95–110 of PrP^C. a, COS-7 cells were transfected with expression plasmids directing the expression of each of the indicated PrP deletion mutants (green, octapeptide repeats; red, globular domain; SS, signal sequence). Transfected cells were assessed for binding of oligomeric Aβ42, or by live cell immunocytochemistry with 6D11 and 7D9 anti-PrP antibodies. Mean and s.e.m. from four experiments. **b**, Schematic of antibody epitopes. **c, d**, PrP^C-expressing COS-7 cells were analysed for oligomeric Aβ42 binding after exposure to the various anti-PrP antibodies for 1 h. 6D11 and 8G8 but not other antibodies block Aβ42-oligomer binding. Data are mean ± s.e.m. from four experiments. Inhibition of binding by 6D11 or 8G8 is significant (*$P < 0.02$, analysis of variance, ANOVA).

a polymorphism in *Prnp* gene is associated with Alzheimer's disease in certain populations[24] and with long-term memory formation in the general population[25]. Amongst several proteins found in PrP^C immunoprecipitates are APP and the related proteins APLP1 and APLP2 (refs 26 and 27).

Glutamate receptors are central to LTP and their modulation has been implicated in deleterious synaptic amyloid-β action[5,8,28,29].
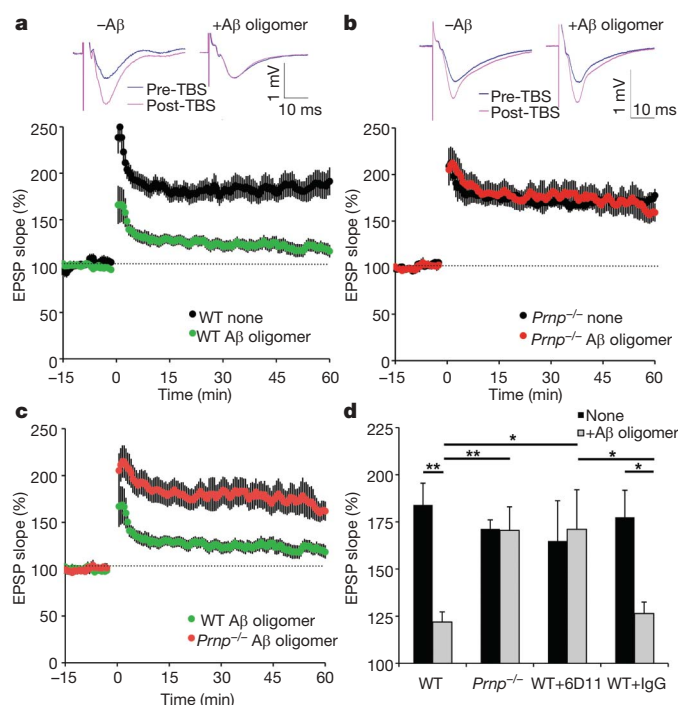
Recently, PrP^C has been shown to interact with *N*-methyl-D-aspartate receptor subunit 2D (NR2D, also known as GRIN2D), and to modulate its function[30]. We assessed whether Aβ42 interaction might regulate glutamate receptors directly through PrP^C. When expressed in a heterologous *Xenopus laevis* oocyte system, GluR1–4 (GRIA1–4) receptors and NR2B (GRIN2B)- and NR2D-containing receptors are insensitive to Aβ42 oligomers, with or without PrP^C co-expression (Supplementary Figs 14 and 15). This is consistent with previous observations that amyloid-β drives glutamate receptor redistribution in neurons together with morphologic changes in dendrites[5,8,9]. Thus,

Figure 4 | PrP^C is required for Aβ42 oligomer inhibition of hippocampal LTP. a, Field potentials were recorded from the CA1 region of hippocampal slices from adult wild-type mice with or without the addition of 500 nM oligomeric Aβ42 to the perfusion 20–40 min before theta burst stimulation (TBS). The top panels show traces before and after TBS. The slope of the EPSP relative to the pre-TBS level is a plotted as a function of time in the lower panel. Data are mean ± s.e.m. from separate slices. For no peptide, $n = 12$ slices from 9 mice; for amyloid-β oligomer, $n = 31$ slices from 14 mice. **b**, CA1 potentials were recorded from slices of mice lacking PrP expression by the same method as in **a**. There is no significant inhibition of LTP by oligomeric Aβ42. For no peptide, $n = 10$ slices from 7 mice; for amyloid-β oligomer, $n = 35$ slices from 15 mice. **c**, The CA1 EPSP slope in wild-type and *Prnp^−/−* slices was recorded in the presence of oligomeric Aβ42 by an observer blind to the genotype and is replotted from panels **a** and **b**. For the values 30–60 min post-TBS, the EPSPs were significantly greater in the *Prnp^−/−* slices by repeated measures ANOVA, $P = 0.005$. For WT, $n = 31$ slices from 14 mice; for *Prnp^−/−*, $n = 35$ slices from 15 mice. **d**, The magnitude of LTP between 30 min and 60 min is plotted as a function of genotype, the addition of 6D11 antibody, control immunoglobulin G (IgG) and/or Aβ42 oligomer before the induction of LTP. Data are mean and s.e.m. For 6D11 without amyloid-β, $n = 7$; for 6D11 plus amyloid-β, $n = 6$. For IgG without amyloid-β, $n = 8$; for IgG plus amyloid-β, $n = 6$. The indicated comparisons are significant at **$P < 0.01$ or *$P < 0.05$, ANOVA. Untreated, IgG and 6D11 slices without Aβ42 did not differ significantly.

1131

the Aβ42-oligomer interaction with PrP[C] is likely to initiate a signalling cascade that is not operative in oocytes, but one that is capable of modifying synaptic morphology and function in the brain. The mechanism by which Aβ42-oligomer binding to PrP[C] participates in Alzheimer's disease appears unrelated to the infectious PrP[Sc] conformation of PrP. In this regard, the neurodegeneration reported in transgenic mice expressing truncated forms of PrP[C] may be more relevant[17,18]. A putative PrP[C]-associated transmembrane co-receptor is likely to have a central role in Alzheimer's-disease-mediated neurodegeneration. PrP[C]-specific reagents will provide molecular tools to dissect the cellular basis for Aβ42-oligomer-induced changes in synaptic function. The interaction between amyloid-β and PrP[C] provides a new site for the development of therapeutics designed to relieve Alzheimer's disease symptoms.

METHODS SUMMARY

Mouse strains. $Prnp^{-/-}$ mice (Edinburgh strain)[21] on an inbred C57Bl6 background were obtained from B. Chesebro and $Prnp^{-/-}$ mice (Zurich I)[20] on a mixed strain background were obtained from the European Mutant Mouse Archive.

Aβ42 preparation and cellular binding. Aβ42 oligomer preparations were generated from synthetic peptide[4]. For binding assays, COS-7 cells were transiently transfected with cDNA expression plasmids or isolated hippocampal neurons were cultured from E18 embryos. Bound biotin–Aβ42 was detected using avidin conjugates.

Electrophysiology. Hippocampal slices (400 μm) from C57Bl6J or $Prnp^{-/-}$ mice were bathed in oxygenated artificial cerebrospinal fluid. The Schaffer collateral pathway was stimulated at 0.033 Hz at levels that evoked less than 50% of maximal field EPSPs. Evoked CA1 field potentials were recorded and the slope of the EPSP determined (Clampfit, Molecular Devices). Aβ42 or antibodies were bath-applied for 20–40 min before inducing LTP with ten 100-Hz trains at five pulses delivered at 5 Hz.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353–356 (2002).
2. Lesne, S. *et al.* A specific amyloid-β protein assembly in the brain impairs memory. *Nature* **440**, 352–357 (2006).
3. Cleary, J. P. *et al.* Natural oligomers of the amyloid-β protein specifically disrupt cognitive function. *Nature Neurosci.* **8**, 79–84 (2005).
4. Chromy, B. A. *et al.* Self-assembly of Aβ$_{1-42}$ into globular neurotoxins. *Biochemistry* **42**, 12749–12760 (2003).
5. Lacor, P. N. *et al.* Abeta oligomer-induced aberrations in synapse composition, shape, and density provide a molecular basis for loss of connectivity in Alzheimer's disease. *J. Neurosci.* **27**, 796–807 (2007).
6. Lacor, P. N. *et al.* Synaptic targeting by Alzheimer's-related amyloid β oligomers. *J. Neurosci.* **24**, 10191–10200 (2004).
7. Walsh, D. M. *et al.* Naturally secreted oligomers of amyloid β protein potently inhibit hippocampal long-term potentiation *in vivo*. *Nature* **416**, 535–539 (2002).
8. Shankar, G. M. *et al.* Natural oligomers of the Alzheimer amyloid-β protein induce reversible synapse loss by modulating an NMDA-type glutamate receptor-dependent signaling pathway. *J. Neurosci.* **27**, 2866–2875 (2007).
9. Shankar, G. M. *et al.* Amyloid-β protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nature Med.* **14**, 837–842 (2008).
10. Hepler, R. W. *et al.* Solution state characterization of amyloid β-derived diffusible ligands. *Biochemistry* **45**, 15157–15167 (2006).
11. Lambert, M. P. *et al.* Diffusible, nonfibrillar ligands derived from Aβ$_{1-42}$ are potent central nervous system neurotoxins. *Proc. Natl Acad. Sci. USA* **95**, 6448–6453 (1998).
12. Prusiner, S. B. Prions. *Proc. Natl Acad. Sci. USA* **95**, 13363–13383 (1998).
13. Yan, S. D. *et al.* RAGE and amyloid-β peptide neurotoxicity in Alzheimer's disease. *Nature* **382**, 685–691 (1996).
14. Wang, H. Y. *et al.* β-Amyloid$_{1-42}$ binds to α7 nicotinic acetylcholine receptor with high affinity. Implications for Alzheimer's disease pathology. *J. Biol. Chem.* **275**, 5626–5632 (2000).
15. Viles, J. H. *et al.* Copper binding to the prion protein: structural implications of four identical cooperative binding sites. *Proc. Natl Acad. Sci. USA* **96**, 2042–2047 (1999).
16. Jackson, G. S. *et al.* Location and properties of metal-binding sites on the human prion protein. *Proc. Natl Acad. Sci. USA* **98**, 8531–8535 (2001).
17. Baumann, F. *et al.* Lethal recessive myelin toxicity of prion protein lacking its central domain. *EMBO J.* **26**, 538–547 (2007).
18. Li, A. *et al.* Neonatal lethality in transgenic mice expressing prion protein with a deletion of residues 105–125. *EMBO J.* **26**, 548–558 (2007).
19. Riek, R. *et al.* NMR structure of the mouse prion protein domain PrP(121–321). *Nature* **382**, 180–182 (1996).
20. Bueler, H. *et al.* Normal development and behaviour of mice lacking the neuronal cell-surface PrP protein. *Nature* **356**, 577–582 (1992).
21. Manson, J. C. *et al.* 129/Ola mice carrying a null mutation in PrP that abolishes mRNA production are developmentally normal. *Mol. Neurobiol.* **8**, 121–127 (1994).
22. Lledo, P. M., Tremblay, P., DeArmond, S. J., Prusiner, S. B. & Nicoll, R. A. Mice deficient for prion protein exhibit normal neuronal excitability and synaptic transmission in the hippocampus. *Proc. Natl Acad. Sci. USA* **93**, 2403–2407 (1996).
23. Curtis, J., Errington, M., Bliss, T., Voss, K. & MacLeod, N. Age-dependent loss of PTP and LTP in the hippocampus of PrP-null mice. *Neurobiol. Dis.* **13**, 55–62 (2003).
24. Riemenschneider, M. *et al.* Prion protein codon 129 polymorphism and risk of Alzheimer disease. *Neurology* **63**, 364–366 (2004).
25. Papassotiropoulos, A. *et al.* The prion gene is associated with human long-term memory. *Hum. Mol. Genet.* **14**, 2241–2246 (2005).
26. Yehiely, F. *et al.* Identification of candidate proteins binding to prion protein. *Neurobiol. Dis.* **3**, 339–355 (1997).
27. Schmitt-Ulms, G. *et al.* Time-controlled transcardiac perfusion cross-linking for the study of protein interactions in complex tissues. *Nature Biotechnol.* **22**, 724–731 (2004).
28. Venkitaramani, D. V. *et al.* β-amyloid modulation of synaptic transmission and plasticity. *J. Neurosci.* **27**, 11832–11837 (2007).
29. Hsieh, H. *et al.* AMPAR removal underlies Aβ-induced synaptic depression and dendritic spine loss. *Neuron* **52**, 831–843 (2006).
30. Khosravani, H. *et al.* Prion protein attenuates excitotoxicity by inhibiting NMDA receptors. *J. Cell Biol.* **181**, 551–565 (2008).

Author Contributions J.L. performed the amyloid-β binding and expression cloning experiments, D.A.G. conducted mouse breeding and tissue biochemistry, S.M.S. and H.B.N. performed the hippocampal electrophysiology experiments, and S.M.S., J.W.G. and J.L. performed the *X. laevis* studies. S.M.S. supervised all experiments. All authors participated in writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.M.S. (stephen.strittmatter@yale.edu).

METHODS

Aβ42 preparations. Oligomerized human Aβ42 was prepared as described[5,6] with minor modifications. Aβ42 peptide or Aβ42 with a biotin moiety attached by 6-carbon linker to the N terminus (both synthesized at Yale in the Keck Biotechnology Resource Laboratory, ~99% pure) were dissolved in 1,1,1,3,3,3-hexafluoro-2-propanol, dried and stored at −80 °C. The dried peptide was reconstituted in anhydrous DMSO at 100 mM, and then diluted in F12 media (Invitrogen) to a final concentration of 100 μM. After 16 h incubation at 22 °C, the preparation was centrifuged at 21,000*g* for 15 min, and the supernatant was collected for experiments.

Mouse strains. $Prnp^{-/-}$ mice (Edinburgh strain)[21] on an inbred C57Bl6 background (greater than ten backcrosses) were obtained from B. Chesebro, and $Prnp^{-/-}$ mice (Zurich I)[20] on a mixed strain background were obtained from the European Mutant Mouse Archive. All electrophysiological strain comparisons were completed by experimentalists unaware of the mouse genotype.

Biophysical characterization of Aβ42 oligomers. For electron microscopic analysis, oligomerized human Aβ42 was prepared as above. Fresh Aβ42 was prepared by dissolving dried peptide similarly into DMSO and then into F12 media immediately before use. Aβ42 fibrils were prepared by dissolving dried peptide in DMSO and then diluting into PBS followed by 24 h incubation at 37 °C with 100 r.p.m. mixing. All Aβ42 preparations were stained with 2% phosphotungstic acid (pH 7.0) for 2 min and imaged with Tecnai 12 Biotwin transmission electron microscope at 160,000-fold magnification. For analysis of the molecular weight of oligomerized human Aβ42, SEC combined with light scattering analysis was used. The SEC data were collected using a sequential connection of Superdex 200, Superdex 75, and Superdex peptide, 10/30, HR SEC columns (GE Healthcare), connected to high-performance liquid chromatography system, Alliance 2965 (Waters Corp.), equipped with an autosampler. The elution from SEC was monitored by a photodiode array (PDA) UV/VIS detector (996 PDA, Waters Corp.), differential refractometer (OPTI-rEx, Wyatt Corp.) and static, multiangle laser LS detector (DAWN-EOS, Wyatt Corp.). The SEC-UV/LS/RI system was equilibrated in F12 media at the flow rate of 0.5 ml min^{-1}. Two software packages were used for data collection and analysis: the Millennium software (Waters Corp.) controlled the high-performance liquid chromatography operation and data collection from the multi-wavelength UV/VIS detector, whereas the ASTRA software (Wyatt Corp.) collected data from the light scattering detector, and recorded the UV trace at 220 nm sent from the PDA detector. The weight average molecular masses were determined across the entire elution profile in the intervals of 1 s from the static LS measurement using the ASTRA software as described previously[31].

Expression cloning and COS-7 cell binding assays. We screened an arrayed adult mouse brain cDNA library (Origene) of 225,000 clones by transfecting COS-7 cells with pools of 5,000 clones, and visually inspecting for those rare cells transfected with a cDNA conferring Aβ42-oligomer binding. Two days after transfection, the cells were incubated with Aβ42 oligomers (1 μM total Aβ42 monomer concentration) in F12 media at 22 °C for 2 h. Next, unbound Aβ42 was removed by extensive washing with F12 media. The cells were fixed with 3.7% formaldehyde, washed three times with PBS, incubated in 65 °C for 2 h, blocked for 20 min with 3% goat serum and 0.1% Triton X-100 in PBS, and incubated for 16 h with alkaline-phosphatase-conjugated steptavidin in PBS supplemented with 1.5% goat serum and 0.05% Triton X-100. Finally, bound alkaline phosphatase was visualized by 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium reaction. The initial screen was followed by sequential fractionation of sub-pools. This process led to identification of two independent cDNA clones mediating high-affinity binding of Aβ42 oligomers to transfected COS-7 cells. Sequencing of the inserts revealed that both clones encode full-length PrP^C. We also performed a directed experiment to identify putative low-affinity Aβ42-oligomer receptor(s): a library consisting of 352 individual preparations of cDNAs encoding transmembrane proteins (GFC-transfection array panel, Origene) was transfected into COS-7 cells, and binding was detected as above. We also tested similarly several candidate receptors; the cDNA plasmids were obtained from Origene or Open Biosystems. Quantification of bound Aβ42 to PrP-transfected COS-7 cells was performed with ImageJ similarly to that described previously[32–35].

For analysis of oligomeric Aβ42 binding to mouse PrP deletion constructs, the indicated constructs were transfected into COS-7 cells. PrP Δ105–125 expression construct was provided by D. Harris[18]. PrP Δ32–121, PrP Δ32–106 and PrP Δ52–91 were provided by E. Flechsig and C. Weismann[36]. To obtain PrP Δ70–105 and PrP Δ95–105 expression constructs, nucleotides encoding mouse PrP amino acids 106–254 and the C-terminal stop codon were PCR-amplified with primers containing EcoRI and XhoI sites, and the fragment was cloned into EcoRI and XhoI sites of pcDNA3.1/mycHis A (Invitrogen). Next, nucleotides encoding mouse PrP amino acids 1–69 or 1–94 were PCR-amplified with primers containing HindIII

and EcoRI sites, and the fragments were cloned into HindIII and EcoRI sites of the above-mentioned PrP 106–254 pcDNA3.1/mycHis A construct.

For double-immunostaining of PrP^C expressed on COS-7 cells and bound non-biotinylated Aβ42, the rabbit anti-Aβ42 antibody (Cell Signaling Technology) was used in combination with 7D9 PrP mouse monoclonal antibody (Covance). Alexa Fluor 488 goat anti-rabbit and 568 goat anti-mouse secondary antibodies (Invitrogen) were used. No specific signal was obtained with amyloid-β antibody without the addition of amyloid-β on PrP-expressing cells, or when incubating amyloid-β with non-transfected control cells.

Antibody blocking experiments. The following PrP antibodies were used: S8B4 (mouse monoclonal; epitope mapped to amino acids 32–69 of PrP (Supplementary Fig. 8), 8B4 clone from Santa Cruz Biotechnology), M-20 (affinity-purified goat polyclonal; raised against peptide located in the C-terminal half of mouse PrP (Supplementary Fig. 8), Santa Cruz Biotechnology), 7D9 (mouse monoclonal; according to manufacturer epitope between amino acids 105 and 125 of mouse PrP, Covance/Signet), 6D11 (mouse monoclonal; according to manufacturer epitope between amino acids 93 and 109, Covance/Signet), A8B4 (mouse monoclonal; according to manufacturer epitope between amino acids 37 and 44 of mouse PrP, 8B4 clone from Alicon GH), MAB5424 (mouse monoclonal; epitope mapped to amino acids 105–125 (Supplementary Fig. 8), Chemicon), and 8G8 (mouse monoclonal; epitope between amino acids 95 and 110, Cayman Chemical/SPIbio). In live-cell staining application, all antibodies robustly recognized PrP expressed on COS-7 cells (Supplementary Fig. 8). To detect the blockade of Aβ42–PrP interaction, COS-7 cells transfected 2 days earlier with PrP expression vector were pre-incubated with the indicated concentrations of antibodies in F12 media at 22 °C for 1 h. Next, biotin–Aβ42 oligomers were added to a final concentration of 250 nM for 2 h. Finally, bound Aβ42 was visualized as described previously.

PrP–Fc pull-down experiments. HEK293T cells were transiently transfected with expression vector encoding mouse PrP (codon 1–230, devoid of GPI-anchor modification sequence) fused to the Fc region of human IgG1 (ref. 37). Secreted PrP–Fc fusion was bound on protein G Sepharose CL-4B (GE Healthcare), and contaminants were removed by washes with increasing concentrations of buffered NaCl solution. Control human IgG Fc fragment (Bethyl Laboratories) was similarly coupled to protein G Sepharose. Aged oligomerized or fresh Aβ42 was incubated with Fc proteins for 2 h at 4 °C, washed extensively with F12, and the bound material was analysed by western blotting with 6E10 amyloid-β antibody. For quantification of bound material, infrared fluorescence secondary antibodies were used in combination with Odyssey infrared imaging system (LI-COR Biosciences); the signal across the whole lane was quantified with Odyssey software. The reported result is average ± s.e.m. from four experiments.

Neuron cultures and neuronal binding assays. Rat and mouse E18 hippocampal neurons were cultured in Neurobasal-A media supplemented with B-27, 0.5 mM L-glutamine, penicillin and streptomycin (all from Invitrogen) on tissue culture plates or on Lab-Tek II-CC2 chamber slides (Nunc) coated with poly-L-lysine and laminin. For the analysis of PrP expression, neurons were lysed with RIPA buffer after the indicated time *in vitro*. Also wild-type and $Prnp^{-/-}$ mouse brain lysates and PrP-expressing HEK293T cell lysates were prepared. The lysates were analysed by western blotting with 8H4 anti-PrP antibody (Alicon AG), anti-βIII-tubulin antibody (Covance), and anti-GAPDH antibody (Sigma). To remove carbohydrate-moieties from PrP, some samples were treated with PNGase F (New England Biolabs). For immunocytochemistry, the following antibodies were used: anti-PrP (mouse monoclonal, 8H4 clone, Alicon AG), anti-MAP2 (chicken polyclonal, Chemicon) and anti-PSD95 (rabbit polyclonal, Zymed/Invitrogen). For rat neurons, binding of Aβ42 was determined as in COS-7 cell binding assays. To determine bound biotin Aβ42 oligomers in wild-type and $Prnp^{-/-}$ cultures, an Aβ42-oligomer concentration equivalent to 80 nM of monomer Aβ42 (~0.3 nM oligomer) was used. The bound biotin–Aβ42-oligomers were visualized with NeutrAvidin Oregon Green 488 conjugate (Invitrogen) and dendrites with anti-MAP2 antibody (chicken polyclonal, Chemicon). Fluorescent images were captured at ×63 with fixed illumination and exposures, and then processed using NIH Image. Integrated signal intensity for bound Aβ42 oligomer in puncta of less than 50 μm^2 was measured after thresholding and background subtraction across both wild-type and $Prnp^{-/-}$ images. The integrated punctate amyloid-β signal was divided be the area of MAP-2-positive dendrite in each image, to generate units of Aβ42 oligomer binding to puncta per dendritic area.

Electrophysiology. Adult (2–6 months of age) C57BL/6J or $Prnp^{-/-}$ mice were decapitated, and their brains rapidly removed and immersed in ice-cold artificial cerebrospinal fluid (aCSF). The composition of the aCSF was as follows: 119 mM NaCl, 2.5 mM KCl, 1.3 mM MgCl$_2$, 2.4 mM CaCl$_2$, 26.2 mM NaHCO$_3$, 11 mM D-glucose, 1.25 mM NaH$_2$PO$_4$. Coronal sections (400 μm) were prepared on a Vibratome 1000 Plus, using stainless steel razor blades. Slices were allowed to recover for at least 2 h in a submerged incubation chamber (BSC-PC, Warner

Instruments) at room temperature, continuously bubbled with a mixture of 95% O_2 and 5% CO_2.

For extracellular recordings, slices were submerged in a recording chamber (RC-27L, Warner Instruments), continuously perfused with oxygenated aCSF at a rate of 2 ml min^{-1} at 30 °C. A bipolar tungsten microelectrode (TM33CCNON, World Precision Instruments) was used to stimulate Schaffer collateral fibres, and extracellular field EPSPs were recorded using a glass microelectrode (2–6 MΩ) filled with aCSF in the stratum radiatum of CA1. For all experiments, test stimuli were given at 0.033 Hz, and the stimulus intensity was set to give baseline field EPSP slopes of less than 50% of the maximal response. A stable baseline was recorded for at least 20 min before inducing LTP. LTP was induced by theta burst stimulation (10 bursts of 4 shocks at 100 Hz, with an interburst interval of 200 ms) given at baseline intensity. Field potentials were recorded using an Axon Intruments 700B amplifier and a Digidata 1440A digitizer, and data were analysed using pClamp 10 software (Molecular Devices).

Experiments with $Prnp^{-/-}$ mice were conducted in a blinded fashion with respect to genotype. Aβ42 oligomers (500 nM), 6D11 anti-PrP^C antibody (5 μg ml^{-1}) and mouse IgG (Sigma-Aldrich, 5 μg ml^{-1}) were diluted in aCSF. For experiments using amyloid-β, prion protein antibody or IgG alone, submerged slices were continually perfused with aCSF containing one of these for 20–40 min before LTP induction, and the same aCSF was recycled for the duration of the recording. For experiments using a combination of amyloid-β oligomers with either anti-PrP antibody or IgG, slices were first perfused with aCSF containing antibody or IgG for 20–40 min followed by the addition of aCSF containing amyloid-β oligomers, and perfused for at least another 20 min before LTP induction. The aCSF was recycled, ensuring the presence of amyloid-β oligomers with either antibody or IgG for the duration of the recording.

Similar results were obtained with the Zurich I and the Edinburgh $Prnp^{-/-}$ mice (Supplementary Fig. 13), and the pooled results are presented in Fig. 4.

31. Folta-Stogniew, E. Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. *Methods Mol. Biol.* **328**, 97–112 (2006).
32. Rajagopalan, S. *et al.* Neogenin mediates the action of repulsive guidance molecule. *Nature Cell Biol.* **6**, 756–762 (2004).
33. Fournier, A. E., GrandPre, T. & Strittmatter, S. M. Identification of a receptor mediating Nogo-66 inhibition of axonal regeneration. *Nature* **409**, 341–346 (2001).
34. Takahashi, T. *et al.* Plexin–neuropilin-1 complexes form functional semaphorin-3A receptors. *Cell* **99**, 59–69 (1999).
35. Takahashi, T., Nakamura, F., Jin, Z., Kalb, R. G. & Strittmatter, S. M. Semaphorins A and E act as antagonists of neuropilin-1 and agonists of neuropilin-2 receptors. *Nature Neurosci.* **1**, 487–493 (1998).
36. Weissmann, C. & Flechsig, E. PrP knock-out and PrP transgenic mice in prion research. *Br. Med. Bull.* **66**, 43–60 (2003).
37. Chen, S., Mange, A., Dong, L., Lehmann, S. & Schachner, M. Prion protein as trans-interacting partner for neurons is involved in neurite outgrowth and neuronal survival. *Mol. Cell. Neurosci.* **22**, 227–233 (2003).

LETTERS

Intracortical circuits of pyramidal neurons reflect their long-range axonal targets

Solange P. Brown[1] & Shaul Hestrin[1]

Cortical columns generate separate streams of information that are distributed to numerous cortical and subcortical brain regions[1]. We asked whether local intracortical circuits reflect these different processing streams by testing whether the intracortical connectivity among pyramidal neurons reflects their long-range axonal targets. We recorded simultaneously from up to four retrogradely labelled pyramidal neurons that projected to the superior colliculus, the contralateral striatum or the contralateral cortex to assess their synaptic connectivity. Here we show that the probability of synaptic connection depends on the functional identities of both the presynaptic and postsynaptic neurons. We first found that the frequency of monosynaptic connections among corticostriatal pyramidal neurons is significantly higher than among corticocortical or corticotectal pyramidal neurons. We then show that the probability of feed-forward connections from corticocortical neurons to corticotectal neurons is approximately three- to four-fold higher than the probability of monosynaptic connections among corticocortical or corticotectal cells. Moreover, we found that the average axodendritic overlap of the presynaptic and postsynaptic pyramidal neurons could not fully explain the differences in connection probability that we observed. The selective synaptic interactions we describe demonstrate that the organization of local networks of pyramidal cells reflects the long-range targets of both the presynaptic and postsynaptic neurons.

The long-range axonal projections of cortical pyramidal neurons target unique sets of cortical and subcortical brain regions and define different functional classes of pyramidal neuron[1]. In addition, each pyramidal neuron elaborates extensive intracortical axon collaterals that generate the majority of excitatory input in neighbouring cortical neurons[2–4]. Recent work has shown that the probability of connection among pyramidal neurons is not homogeneous[5–11]. However, whether local synaptic interactions reflect the long-range axonal projections of both the presynaptic and the postsynaptic partner is not known.

Non-overlapping populations of pyramidal neurons projecting to different brain regions are intermingled within layer 5 (L5), the main output layer of the cortex[1]. We first compared the homotypic connectivity among L5 pyramidal neurons projecting to different brain regions. To address this question, we injected fluorescent latex microspheres into the ipsilateral superior colliculus to label corticotectal neurons, the contralateral striatum to label corticostriatal neurons, or the contralateral visual cortex to label corticocortical neurons. We next recorded in whole cell configuration from fluorescently labelled neurons and determined that the intrinsic physiological properties of corticotectal, corticostriatal and corticocortical pyramidal neurons were significantly different (Supplementary Fig. 1, Supplementary Table 1), as expected for three distinct classes of pyramidal neuron[12–14].

To assay the synaptic connectivity among pyramidal neurons projecting to the same long-range target, we recorded simultaneously

from multiple fluorescently labelled neurons using whole-cell patch clamp techniques. Action potentials were generated with brief current injections in each neuron in turn while the synaptic responses in the other neurons were recorded. In synaptically connected cells, these presynaptic action potentials elicited monosynaptic unitary excitatory postsynaptic potentials (EPSPs) in the postsynaptic partner. Monosynaptic connections were identified between neurons for all three cell types (Fig. 1a–c). The synaptic properties, including the mean amplitudes and the paired-pulse ratio, were similar among the three types of connection (Supplementary Table 2).

Although the properties of the synaptic responses were similar, the rate of monosynaptic connections among corticostriatal neurons was
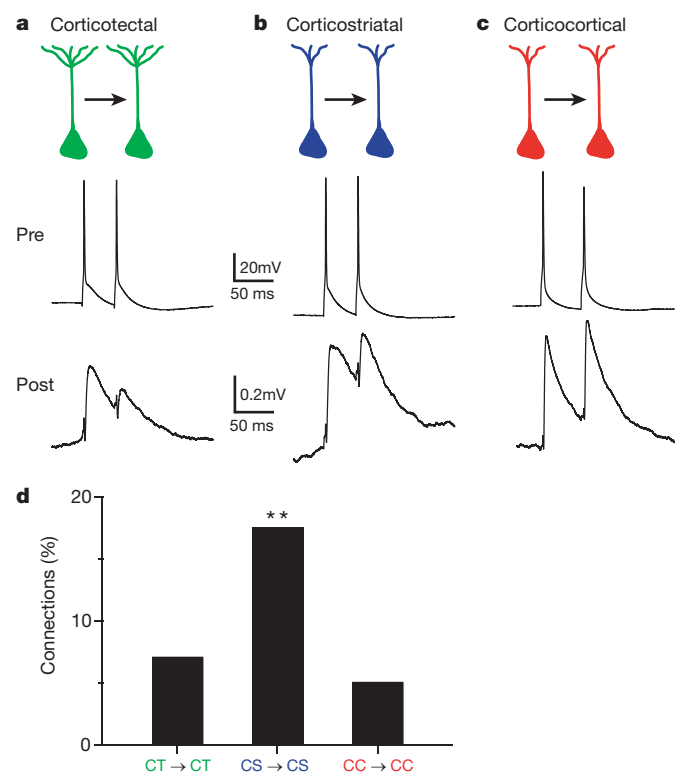

Figure 1 | Different frequencies of monosynaptic connections between corticotectal, corticostriatal or corticocortical neurons. a, **b**, **c**, Presynaptic action potentials elicit a synaptic response in a postsynaptic cell during simultaneous recordings from two monosynaptically connected corticotectal neurons (**a**), corticostriatal neurons (**b**) and corticocortical neurons (**c**). **d**, The frequency of identified monosynaptic connections among connections tested is shown for corticotectal (CT) connections, corticostriatal (CS) connections and corticocortical (CC) connections. **P < 0.05.

[1]Department of Comparative Medicine, Stanford University School of Medicine, 300 Pasteur Drive, Edwards Building, R314, Stanford, California 94305-5342, USA.

significantly higher than the rate among corticotectal neurons or among corticocortical neurons. Eighteen per cent of corticostriatal-to-corticostriatal potential connections tested were monosynaptically connected (seven of 40 tested connections), a considerably higher connectivity than previously reported for L5 pyramidal neurons[6,10,11,15,16]. By contrast, only 7% of potential corticotectal-to-corticotectal connections (16 of 225 tested connections) and 5% of potential corticocortical-to-corticocortical connections (6 of 118 tested connections) were monosynaptically connected ($P = 0.034$, Pearson's chi-squared test; Fig. 1d). Our results indicate that specific functional classes of pyramidal neuron can form highly interconnected networks embedded within the local circuitry of the cortex.

Several connectivity schemes could underlie the observed differences in the probability of connection among these cell types. First, the corticotectal and corticocortical cells we studied were located in the visual cortex whereas the corticostriatal cells were located in the sensorimotor cortex, raising the possibility of a regional effect on cortical connectivity. Second, each presynaptic cell type could connect to its neighbours with a characteristic frequency. Corticostriatal neurons may simply connect to all their targets with a higher probability than corticotectal or corticocortical neurons. This interpretation is consistent with recent work showing that pyramidal neurons with different long-range projections have different probabilities of forming connections with neighbouring neurons[10,11,17], suggesting that, rather than reflecting axonal target selectivity *per se*, the probability of connection is a global property specific to each pyramidal cell type. Third, cortical circuits may reflect the functional identity of the presynaptic and postsynaptic cell types. In this case, the intracortical connectivity among pyramidal neurons may reflect the long-range axonal projections of both the presynaptic and postsynaptic pyramidal neurons. Whether pyramidal neurons can synapse differentially onto neighbouring pyramidal neurons of different functional classes is not known.

To differentiate among these possibilities, we targeted quadruplets composed of pyramidal neurons with two different long-range projections for electrophysiological recordings. This configuration allowed us to simultaneously compare the connectivity rates of two types of pyramidal neuron with two different postsynaptic targets. If the brain region or the functional identity of the presynaptic neuron dictates its connectivity with neighbouring pyramidal neurons, we expect that a pyramidal neuron's probabilities of connection with the two different postsynaptic targets are the same. However, if the probability of connection differs for the different types of connection, then intracortical connectivity depends on the functional identity of both the presynaptic neuron and the postsynaptic neuron.

We injected red fluorescent microspheres into the contralateral visual cortex and green fluorescent microspheres into the ipsilateral superior colliculus to label both corticocortical and corticotectal neurons in the same animal. We then recorded simultaneously from these classes of pyramidal neuron intermingled in L5 of the visual cortex, and directly compared the probability of connection for four types of connections: corticocortical to corticocortical, corticocortical to corticotectal, corticotectal to corticotectal, and corticotectal to corticocortical (Fig. 2a). We found that the probability of identifying corticotectal-to-corticocortical connections is 5% (4 of 86 connections tested), similar to the 7% probability of identifying a corticotectal-to-corticotectal connection ($P = 0.43$, Pearson's chi-squared test), indicating that corticotectal cells do not connect to corticocortical pyramidal neurons in preference to corticotectal pyramidal neurons. However, the probability of identifying a corticocortical-to-corticotectal connection is 19% (16 of 86 connections tested), whereas the probability of identifying a corticocortical-to-corticocortical connection is only 5% ($P = 0.002$, Pearson's chi-squared test; Fig. 2b), indicating that corticocortical pyramidal neurons preferentially target neighbouring corticotectal neurons.

Our results indicate that the probability of identifying connections between L5 pyramidal neurons in visual cortex is not universally low.
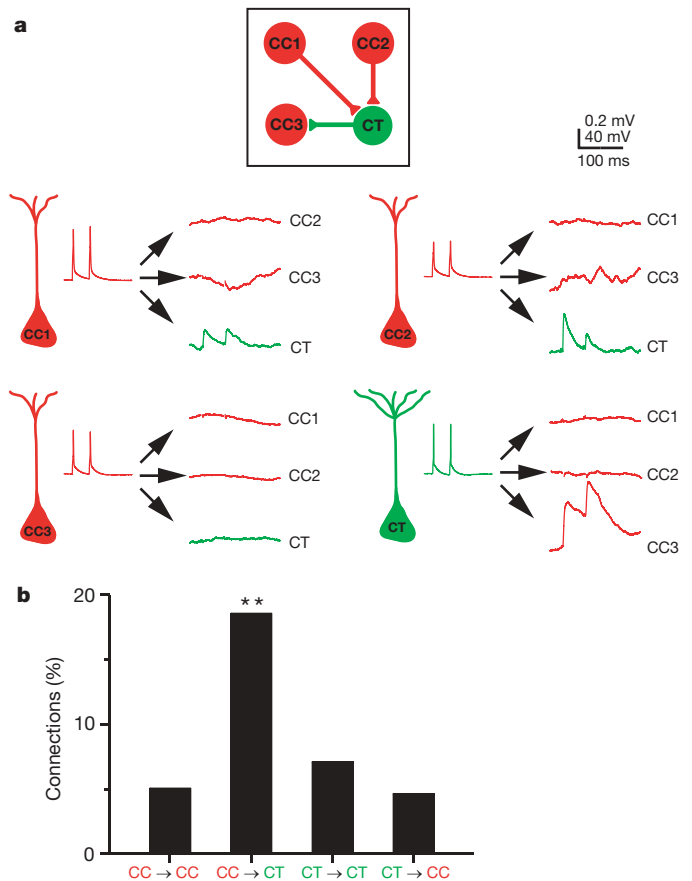

Figure 2 | The probability of connection depends on the identities of the presynaptic and the postsynaptic pyramidal cell types. a, An example of a quadruple recording. Three corticocortical neurons (CC1, CC2, CC3; red) and one corticotectal neuron (green) were recorded simultaneously and the 12 possible synaptic connections were tested. The corticocortical neurons CC1 and CC2 synapsed onto the neighbouring corticotectal neuron. The corticotectal neuron in turn synapsed onto the corticocortical neuron CC3. Scale bars: 40 mV for presynaptic action potentials, 0.2 mV for postsynaptic responses. **b**, The frequency of synaptic connections identified among the tested connections is shown for the four possible types of connection among corticocortical and corticotectal neurons. **$P < 0.01$.

The probability of identifying corticocortical-to-corticotectal connections was as high as the probability of identifying corticostriatal-to-corticostriatal connections in the sensorimotor cortex. Our results further indicate that the connectivity among pyramidal neurons is not simply a global characteristic of the presynaptic or the postsynaptic neuron. A corticocortical axon is almost four times more likely to form a functional synapse with a local corticotectal pyramidal neuron than with a corticocortical pyramidal neuron, indicating that local intracortical circuits reflect the functional identity of the postsynaptic pyramidal neuron. Moreover, the probability of identifying a monosynaptic connection is 19% for corticocortical-to-corticotectal combinations but only 7% for corticotectal-to-corticotectal combinations, indicating that the long-range target of the presynaptic cell is also important. Combined, our results suggest that it is the interplay between the functional identities of the presynaptic and the postsynaptic pyramidal neurons that determines the pattern of local microcircuits in the cortex.

Several authors have suggested that pyramidal neurons synapse probabilistically onto neighbouring neurons, and that their connectivity is a function of the average spatial overlap of their dendritic and axonal processes[4,18–21]. If this is the case, the connectivity rates that we measured may simply reflect different average spatial overlaps for the five connections we tested rather than any local selection among different functional types of pyramidal neuron. To evaluate

this possibility, we asked whether the frequencies of monosynaptic connections we measured could be explained by differences in the distribution of the axonal and dendritic processes for these cell types. To address this question, we first reconstructed the three-dimensional morphology of L5 pyramidal neurons of each type filled with biocytin during our physiological recordings. The reconstructions of the dendritic and axonal arbors are shown (in blue and red, respectively) in Fig. 3a–c. For each of the three cell types, the morphology of the L5 reconstructed neurons was similar[11,12,22–24], indicating that the intracortical morphology of each functional class was consistent. However, the distribution of the dendritic and axonal processes among the three cells types was clearly different (Supplementary Figs 2, 3, 4).

Next we asked whether these morphological differences could account for the differences in connectivity that we measured physiologically. To estimate the axodendritic overlap, we quantified the average local density of the dendritic and axonal processes for each cell by generating length density maps from the three-dimensional reconstructions for each type of process[25]. We then calculated the product of the axonal length density map and the dendritic length density map for each combination of neurons that we studied electrophysiologically. Figure 3d shows the results for neurons separated by 50 μm, the average distance between neurons in our physiological data set (see Methods Summary). Separating neurons from 0 to 200 μm, the largest distance between neurons in our data set, produced similar results. These axodendritic overlaps estimate the potential numbers of synapses formed between each combination of cell types studied.

To determine whether the differences in axodendritic overlap could account for the functional connectivity that we measured, we next integrated the maps of axodendritic overlap to obtain the overall axodendritic overlap for each type of synaptic connection. The probability of connection and the axodendritic overlap for each of the cell combinations tested are plotted in Fig. 3e. If a doubling in axodendritic overlap results in a doubling in the probability of connection, then the ratio of the axodendritic overlaps for two types of cell pair should be equal to the ratio of their probabilities of connection. However, the ratio of connection probabilities for corticocortical-to-corticotectal connections and corticocortical-to-corticocortical connections, for example, was 3.7, whereas the ratio of axodendritic overlaps was 1.6. The resulting ratio of these two numbers was significantly greater than one ($P = 0.03$). Previous work has shown that synapses among neighbouring L5 pyramidal neurons are largely located on the proximal dendrites[10,11,15,26]. Restricting our analyses to the perisomatic dendrites produced similar results ($P = 0.02$). Taken together, our data preclude a straightforward linear relationship between the average axodendritic overlap and the probability of connection, and suggest that the average local density of axons and dendrites alone cannot explain the differences in the probability of connection.

Previous experiments have suggested that pyramidal neurons within L5 form a sparsely connected network, with probabilities of connection ranging from 1% to 12% (refs 6, 10, 11, 15, 16). Here we show that the probability of identifying monosynaptic connections among corticostriatal pyramidal neurons and feed-forward connections from corticocortical to corticotectal pyramidal neurons is approximately 20% per connection tested (equivalent to ~30–40% per pair tested). The excitatory monosynaptic connections among corticostriatal pyramidal neurons we describe could amplify the activation of interconnected ensembles of corticostriatal neurons, and the resulting coherent activity could depolarize functionally related striatal neurons, consistent with the hypothesis that the activity of many converging corticostriatal axons is required to depolarize postsynaptic striatal neurons[27,28]. We show that, although corticocortical neurons are monosynaptically interconnected at low rates, as are corticotectal neurons, the local intracortical axons of corticocortical cells target corticotectal neurons with high probability. Interestingly, in vivo experiments showed that corticotectal cells were preferentially activated by callosal stimulation and suggested that
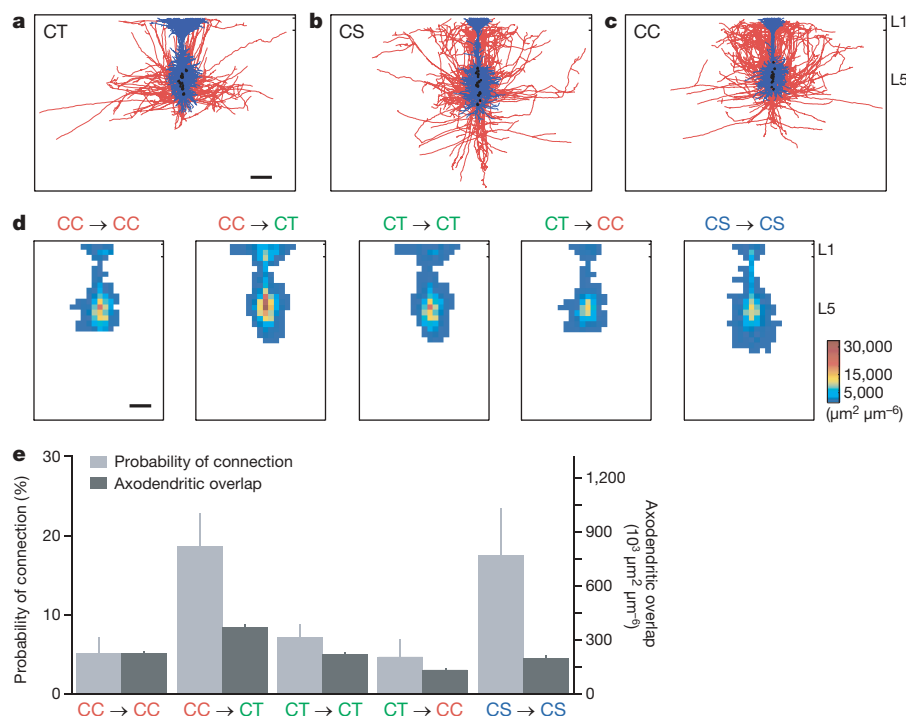

Figure 3 | The average axonal and dendritic architecture alone cannot explain differences in the connection probability. a, b, c, The morphology of 15 corticotectal (**a**), corticostriatal (**b**) and corticocortical neurons (**c**). Blue, dendrites; red, axons; black, somas. **d,** The dendritic and axonal length density maps were used to estimate the spatial overlap of the neuronal processes for the five types of connection tested physiologically. We display the resulting maps of axodendritic overlap (colour scale), generated from cells aligned relative to the pial margin and shifted 50 μm relative to each other. Scale bars, 200 μm. **e,** The probability of physiological connection and the average axodendritic overlap are plotted for each type of connection tested.

feed-forward corticocortical input is important in generating the receptive field properties of corticotectal neurons[29,30]. Our results indicate that the probability of connection among specific functional classes of pyramidal neurons can be quite high, and suggest that highly interconnected functional subnetworks are embedded within the local circuitry of L5.

How local cortical circuits generate the cortex's output remains an open question. Previous work has shown that the connectivity between two pyramidal neurons influences the synaptic input they receive, demonstrating the existence of interconnected subnetworks within the neocortex[6,8,9]. We demonstrate that connections among pyramidal neurons reflect the long-range outputs of both the pre-synaptic and postsynaptic pyramidal neurons. Our results suggest an approach for understanding the function of specialized subnetworks embedded within cortical circuits. Unravelling the local circuits of pyramidal neurons whose long-range targets are known will allow us to understand how the different cortical outputs are generated within the cortical microcircuit. Given the diversity in the distant targets of pyramidal neurons, our findings suggest the existence of multiple networks of pyramidal neurons whose local intracortical connections subserve the specific roles played by their long-range axons.

METHODS SUMMARY

Mice (P14 to P17) were anaesthetized and fluorescently labelled latex micro-spheres (RetroBeads, Lumafluor) were injected into the ipsilateral superior colliculus, the contralateral striatum and the contralateral cortex to retrogradely label cortical neurons projecting to each target. One or more days later, para-sagittal cortical slices were sectioned. Neurons labelled with fluorescent beads were targeted for simultaneous whole-cell patch clamp recordings and their synaptic connectivity was assessed (see Methods). The morphology of the recorded neurons was revealed with biocytin using standard techniques and was reconstructed in three dimensions. To estimate the spatial overlap of the axonal and dendritic processes of the presynaptic and postsynaptic cells, we determined the axonal and dendritic length density maps of each cell. The axonal length density map of each cell of the appropriate presynaptic cell type (either corticotectal, corticostriatal or corticocortical) was multiplied by each dendritic length density map of each cell of the relevant postsynaptic cell type. Because the pairs of cells we studied were separated by an average of $53 \pm 24\,\mu m$ (mean $\pm$ s.d.; $n = 235$ pairs; range, 10–200 μm), we shifted the dendritic length density map 50 μm relative to the axonal length density map to estimate the spatial overlap. These results are compared with the measured physiological connectivity (Fig. 3). Because the distance between pairs of recorded cells ranged from 10 to 200 μm, we also shifted the dendritic length density maps from 0 to 200 μm relative to the axonal length density maps. We performed similar ana-lyses with the neurons aligned by their soma position. Also, we restricted the analysis to the perisomatic dendritic processes, as this is where synapses among L5 pyramidal neurons are largely located[10,11,15,26]. These manipulations all resulted in axodendritic overlaps similar to those shown in Fig. 3 (data not shown).

1. Jones, E. G. in *Cerebral Cortex* (eds Peters, A. & Jones, E. G.) 521–553 (Plenum, 1984).
2. Kisvarday, Z. F. *et al.* Synaptic targets of HRP-filled layer III pyramidal cells in the cat striate cortex. *Exp. Brain Res.* **64,** 541–552 (1986).
3. McGuire, B. A., Hornung, J. P., Gilbert, C. D. & Wiesel, T. N. Patterns of synaptic input to layer 4 of cat striate cortex. *J. Neurosci.* **4,** 3021–3033 (1984).
4. Braitenberg, V. & Schuz, A. *Cortex: Statistics and Geometry of Neuronal Connectivity* (Springer, 1998).
5. Kozloski, J., Hamzei-Sichani, F. & Yuste, R. Stereotyped position of local synaptic targets in neocortex. *Science* **293,** 868–872 (2001).
6. Song, S., Sjostrom, P. J., Reigl, M., Nelson, S. & Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3,** e68 (2005).
7. Wang, Y. *et al.* Heterogeneity in the pyramidal network of the medial prefrontal cortex. *Nature Neurosci.* **9,** 534–542 (2006).
8. Kampa, B. M., Letzkus, J. J. & Stuart, G. J. Cortical feed-forward networks for binding different streams of sensory information. *Nature Neurosci.* **9,** 1472–1473 (2006).
9. Yoshimura, Y., Dantzker, J. L. & Callaway, E. M. Excitatory cortical neurons form fine-scale functional networks. *Nature* **433,** 868–873 (2005).
10. Le Bé, J. V., Silberberg, G., Wang, Y. & Markram, H. Morphological, electrophysiological, and synaptic properties of corticocallosal pyramidal cells in the neonatal rat neocortex. *Cereb. Cortex* **17,** 2204–2213 (2006).
11. Morishima, M. & Kawaguchi, Y. Recurrent connection patterns of corticostriatal pyramidal cells in frontal cortex. *J. Neurosci.* **26,** 4394–4405 (2006).
12. Kasper, E. M., Larkman, A. U., Lubke, J. & Blakemore, C. Pyramidal neurons in layer 5 of the rat visual cortex. I. Correlation among cell morphology, intrinsic electrophysiological properties, and axon targets. *J. Comp. Neurol.* **339,** 459–474 (1994).
13. Wang, Z. & McCormick, D. A. Control of firing mode of corticotectal and corticopontine layer V burst-generating neurons by norepinephrine, acetylcholine, and 1S,3R-ACPD. *J. Neurosci.* **13,** 2199–2216 (1993).
14. Hattox, A. M. & Nelson, S. B. Layer V neurons in mouse cortex projecting to different targets have distinct physiological properties. *J. Neurophysiol.* **98,** 3330–3340 (2007).
15. Markram, H., Lubke, J., Frotscher, M., Roth, A. & Sakmann, B. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol. (Lond.)* **500,** 409–440 (1997).
16. Thomson, A. M., Deuchars, J. & West, D. C. Large, deep layer pyramid-pyramid single axon EPSPs in slices of rat motor cortex display paired pulse and frequency-dependent depression, mediated presynaptically and self-facilitation, mediated postsynaptically. *J. Neurophysiol.* **70,** 2354–2369 (1993).
17. Mercer, A. *et al.* Excitatory connections made by presynaptic cortico-cortical pyramidal cells in layer 6 of the neocortex. *Cereb. Cortex* **15,** 1485–1496 (2005).
18. Hellwig, B. A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. *Biol. Cybern.* **82,** 111–121 (2000).
19. Binzegger, T., Douglas, R. J. & Martin, K. A. A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.* **24,** 8441–8453 (2004).
20. Stepanyants, A. & Chklovskii, D. B. Neurogeometry and potential synaptic connectivity. *Trends Neurosci.* **28,** 387–394 (2005).
21. Kalisman, N., Silberberg, G. & Markram, H. Deriving physical connectivity from neuronal morphology. *Biol. Cybern.* **88,** 210–218 (2003).
22. Hallman, L. E., Schofield, B. R. & Lin, C. S. Dendritic morphology and axon collaterals of corticotectal, corticopontine, and callosal neurons in layer V of primary visual cortex of the hooded rat. *J. Comp. Neurol.* **272,** 149–160 (1988).
23. Hubener, M. & Bolz, J. Morphology of identified projection neurons in layer 5 of rat visual cortex. *Neurosci. Lett.* **94,** 76–81 (1988).
24. Larsen, D. D., Wickersham, I. R. & Callaway, E. M. Retrograde tracing with recombinant rabies virus reveals correlations between projection targets and dendritic architecture in layer 5 of mouse barrel cortex. *Front. Neural Circuits.* doi:10.3389/neuro.04.005.2007 (2008).
25. Shepherd, G. M., Stepanyants, A., Bureau, I., Chklovskii, D. & Svoboda, K. Geometric and functional organization of cortical circuits. *Nature Neurosci.* **8,** 782–790 (2005).
26. Frick, A., Feldmeyer, D., Helmstaedter, M. & Sakmann, B. Monosynaptic connections between pairs of L5A pyramidal neurons in columns of juvenile rat somatosensory cortex. *Cereb. Cortex* **18,** 397–406 (2008).
27. Stern, E. A., Jaeger, D. & Wilson, C. J. Membrane potential synchrony of simultaneously recorded striatal spiny neurons *in vivo. Nature* **394,** 475–478 (1998).
28. Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. & Suarez, H. H. Recurrent excitation in neocortical circuits. *Science* **269,** 981–985 (1995).
29. Swadlow, H. A. Efferent neurons and suspected interneurons in binocular visual cortex of the awake rabbit: receptive fields and binocular properties. *J. Neurophysiol.* **59,** 1162–1187 (1988).
30. Singer, W., Tretter, F. & Cynader, M. Organization of cat striate cortex: a correlation of receptive-field properties with afferent and efferent connections. *J. Neurophysiol.* **38,** 1080–1098 (1975).

Author Contributions S.P.B. and S.H. designed the experiments. S.P.B. collected the data and S.P.B. and S.H. performed the analyses. S.P.B. and S.H. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.H. (shaul.hestrin@stanford.edu).

METHODS

Neuronal labelling. All experimental procedures were approved by the Institutional Animal Care and Use Committee of Stanford University. Juvenile mice (P14 to P17; C57BL/6 x CD-1 and YFP H-line[31]), were anaesthetized and placed in a stereotaxic frame. Using stereotaxic coordinates adjusted for the age of the mice[32], two to 15 sites in the striatum, the superior colliculus and/or the contralateral cortex were injected with 50 nl of a suspension of fluorescently labelled latex microspheres[33] (red or green RetroBeads, Lumafluor). Buprenorphine (0.05 mg kg^{-1}) was administered to alleviate post-operative discomfort. Injections into the superior colliculus labelled neurons in L5 of the ipsilateral visual cortex. Injections into the striatum labelled neurons in layers 2/3 and 5 of ipsilateral and contralateral cortices. Only those cells in L5 of the cortex contralateral to the injection site, representing a subset of corticostriatal cells whose projections include the contralateral striatum, were targeted for further study[34–36]. When studying corticocortical connections, we always simultaneously labelled corticocortical and corticotectal neurons by injecting beads of one colour into the contralateral visual cortex and beads of the other colour in the ipsilateral superior colliculus. There was essentially no overlap between these two cell populations[22,23]. Only those corticocortical cells that intermingled with retrogradely labelled corticotectal pyramidal neurons in L5 were targeted for physiological study. To verify the stereotaxic coordinates of the injections, injected hemispheres were fixed, sectioned and mounted for visualization (Vectashield, Vector Laboratories).

Slice preparation and cell identification. One or more days after the injections, each mouse was anaesthetized and decapitated in an ice-cold sucrose solution composed of 75 mM sucrose, 76 mM NaCl, 25 mM NaHCO$_3$, 25 mM glucose, 2.5 mM KCl, 1.25 mM NaH$_2$PO$_4$, 7 mM MgCl$_2$, 0.5 mM CaCl$_2$, pH 7.4, 325 mosM. Parasagittal cortical slices, 300-µm thick, were sectioned from the selected hemisphere glued on a ramp set at a 30° angle (Integraslice 7550 MM, Campden Instruments), and were maintained in the same solution at 32–34 °C for 30 min before being transferred to artificial cerebrospinal fluid composed of 125 mM NaCl, 2.5 mM KCl, 1.25 mM NaH$_2$PO$_4$, 1 mM MgSO$_4$, 2 mM CaCl$_2$, 26 mM NaHCO$_3$, 20 mM glucose, 4 mM lactic acid, 2 mM pyruvic acid and 0.4 mM ascorbic acid, pH 7.4, 325 mosM, at room temperature (22–25 °C). All solutions were continuously bubbled with 95% O$_2$ and 5% CO$_2$. Retrogradely labelled neurons were identified under epifluorescent illumination (Axioskop 2 FS Plus, ×40 objective, numerical aperture 0.8, Zeiss) and targeted for recording using infrared differential-interference-contrast video microscopy (Sensicam QE, Cooke Corporation).

Electrophysiological recordings. Glass electrodes (2–4 mΩ) were filled with an internal solution containing 2.7 mM KCl, 120 mM potassium methylsulfate, 9 mM HEPES, 0.18 mM EGTA, 4 mM MgATP, 0.3 mM NaGTP, 20 mM phosphocreatine(Na), pH 7.3, 295 mosM. Simultaneous whole-cell patch clamp recordings of the targeted pyramidal cells were obtained using two Multiclamp 700A patch amplifiers (Molecular Devices) in current-clamp mode. All experiments were performed at 32–35 °C. Results were not corrected for the liquid junction potential.

Data acquisition and analysis. All data acquisition and analysis was performed using custom software written in IGOR Pro (Wavemetrics) or Matlab (Mathworks). To compare the adaptation rate of the three cell types, we injected a 200-ms step of depolarizing current adjusted to elicit six to 13 action potentials. A line was fitted to the plot of interspike intervals (ISIs) for each cell (Supplementary Fig. 1d). The first two ISIs were omitted from the analysis because corticotectal cells fired a burst at the start of the current injection. The slope was then divided by the mean ISI to generate an adaptation index for each cell. An adaptation index of zero indicates no adaptation in the spike rate. A positive adaptation index indicates an adapting spike train, whereas a negative adaptation index indicates a spike train with increasingly shorter ISIs. The sag was assessed by fitting a single exponential to the recovery from a hyperpolarizing current step.

Synaptic connectivity was typically assessed by averaging 25 or more traces with two presynaptic action potentials at 20 or 25 Hz and/or 12 presynaptic action potentials at 100 Hz. Each presynaptic action potential was generated by a 3-ms injection of current, and individual trials were separated by 10 s. The 555 potential connections were classified as connected or unconnected while blinded to the identity of the presynaptic and postsynaptic neurons. Recorded neurons were separated by less than 200 µm (mean distance, 53 ± 24 µm (mean ± s.d.); $n = 235$ pairs). There was no significant difference in the distance between pairs of connected neurons and pairs of unconnected neurons for all connection types tested (data not shown). A bias in the vertical position of corticocortical and corticotectal neurons could not account for the differences in connectivity observed. The vertical distance we measured was a positive number when the corticocortical cell was above the corticotectal cell and was negative when the corticocortical cell was below the corticotectal cell. The mean vertical distance was 1 ± 33 µm (mean ± s.d.), which is not significantly different from zero ($P = 0.83$, $n = 86$). There was also no difference in the vertical arrangement of connected and unconnected corticocortical-to-corticotectal pairs ($P = 0.51$).

Morphologic reconstruction and analysis. To reveal the morphology of the recorded neurons, 0.25% w/v biocytin was included in the pipette recording solution of at least one of the pipettes. Following the physiological recordings, the tissue was processed using standard techniques to visualize the neurons with diaminobenzidine. The axons and dendrites of well-stained neurons were reconstructed in three dimensions using a Neurolucida system (Microbrightfield) and a ×100 oil-immersion objective (numerical aperture 1.4, Zeiss). No correction was made for tissue shrinkage.

To analyse the distribution of neuronal processes for each cell, we measured the total length of dendrite or axon for each 50 µm × 50 µm × 300 µm cuboid in a 60 × 40 grid using Neuroexplorer (Microbrightfield). All reconstructed processes were included in this volume. Results from individual cells were then aligned either by soma position or relative to the pial margin. To estimate the spatial overlap of the dendritic and axonal processes of corticotectal, corticostriatal and corticocortical pyramidal neurons, we multiplied each axonal length density map by each dendritic length density map for each combination of cell types tested physiologically, to generate an estimate of the potential number of synapses formed between a pair of neurons. The results were used to compare the potential synaptic connectivity for each combination of cell pairs.

Results are expressed as means plus standard errors unless otherwise noted. The physiological and morphological properties of the three cells types were compared using one-way analysis of variance or the Kruskal–Wallis analysis of variance for multiple comparisons. When only two cell types were compared, Student's t-test was used. The probability of connection was assessed using Pearson's chi-squared test (two tailed). The relationship between the connectivity and the axodendritic overlap was also assessed, using a bootstrap approach to test the null hypothesis that the connectivity and the dendritic overlap were linearly related with a slope of one. The P values ranged from 0.017 to 0.039 using this approach for all the different configurations tested. These configurations included aligning the cell pairs relative to the pial margin or the cell bodies, separating the cell pairs by up to 200 µm, the largest separation in our physiological data set, and including only the perisomatic dendrites in the analysis.

31. Feng, G. et al. Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron* **28**, 41–51 (2000).
32. Paxinos, G. & Fanklin, K. B. J. *The Mouse Brain in Stereotaxic Coordinates* (Academic, 2001).
33. Katz, L. C., Burkhalter, A. & Dreyer, W. J. Fluorescent latex microspheres as a retrograde neuronal marker for in vivo and in vitro studies of visual cortex. *Nature* **310**, 498–500 (1984).
34. Wilson, C. J. Morphology and synaptic connections of crossed corticostriatal neurons in the rat. *J. Comp. Neurol.* **263**, 567–580 (1987).
35. Lei, W., Jiao, Y., Del Mar, N. & Reiner, A. Evidence for differential cortical input to direct pathway versus indirect pathway striatal projection neurons in rats. *J. Neurosci.* **24**, 8289–8299 (2004).
36. Reiner, A., Jiao, Y., Del Mar, N., Laverghetta, A. V. & Lei, W. L. Differential morphology of pyramidal tract-type and intratelencephalically projecting-type corticostriatal neurons and their intrastriatal terminals in rats. *J. Comp. Neurol.* **457**, 420–440 (2003).

Dendritic encoding of sensory stimuli controlled by deep cortical interneurons

Masanori Murayama[1], Enrique Pérez-Garci[1], Thomas Nevian[1], Tobias Bock[1], Walter Senn[1] & Matthew E. Larkum[1]

The computational power of single neurons is greatly enhanced by active dendritic conductances[1] that have a large influence on their spike activity[2–4]. In cortical output neurons such as the large pyramidal cells of layer 5 (L5), activation of apical dendritic calcium channels leads to plateau potentials that increase the gain of the input/output function[5] and switch the cell to burst-firing mode[6–9]. The apical dendrites are innervated by local excitatory and inhibitory inputs as well as thalamic[10–13] and corticocortical projections[14–16], which makes it a formidable task to predict how these inputs influence active dendritic properties *in vivo*. Here we investigate activity in populations of L5 pyramidal dendrites of the somatosensory cortex in awake and anaesthetized rats following sensory stimulation using a new fibre-optic method[17] for recording dendritic calcium changes. We show that the strength of sensory stimulation is encoded in the combined dendritic calcium response of a local population of L5 pyramidal cells in a graded manner. The slope of the stimulus–response function was under the control of a particular subset of inhibitory neurons activated by synaptic inputs predominantly in L5. Recordings from single apical tuft dendrites *in vitro* showed that activity in L5 pyramidal neurons disynaptically coupled via interneurons directly blocks the initiation of dendritic calcium spikes in neighbouring pyramidal neurons. The results constitute a functional description of a cortical microcircuit in awake animals that relies on the active properties of L5 pyramidal dendrites and their very high sensitivity to inhibition. The microcircuit is organized so that local populations of apical dendrites can adaptively encode bottom-up sensory stimuli linearly across their full dynamic range.

To explore how dendritic activity encodes sensory input *in vivo*, we used a microendoscopic technique (the 'periscope' method[17]) ideal for recording from local populations of L5 pyramidal dendrites in anaesthetized and awake animals (Fig. 1a, Methods). This method provides calcium fluorescence signals exclusively from pyramidal apical tuft dendrites by combining two approaches: specific loading of L5 pyramidal dendrites and horizontal imaging of the upper three layers of cortex (Fig. 1a). Fluorescence responses to sensory stimuli were recorded from populations of L5 bolus-loaded dendrites that were not detectable using two-photon imaging from individual dendrites[18] ($n = 6$, data not shown; see also Methods). This method has a high sensitivity to dendritic activity and can be easily applied to awake, freely moving animals.

Air-puff stimulation of the hindlimb under urethane anaesthesia produced biphasic responses in dendrites of the contralateral hindlimb area of the primary somatosensory cortex (Fig. 1b, black). This dendritic population response could be blocked completely by the application of the Ca^{2+} channel blocker Cd^{2+} to the cortical surface (Fig. 1b, grey), leaving no detectable movement artefacts. Application of the $GABA_A$ (GABA, γ-aminobutyric acid) receptor antagonist, gabazine, to the cortical surface increased the first component of the dendritic

population response approximately fivefold ($0.42 \pm 16\%$ versus $2.12 \pm 0.60\%$ (all data given as mean $\pm$ s.d.), $P < 0.05$, $n = 5$; Fig. 1b, c, red), whereas the $GABA_B$ receptor antagonist CGP52432 did not change this component ($0.42 \pm 0.16\%$ versus $0.38 \pm 0.07\%$, $P = 0.29$, $n = 5$; Fig. 1b, c, green). Throughout this study we focused on the factors influencing the increase in the initial component of calcium influx into the dendrites modulated by $GABA_A$ receptors.

We investigated the dendritic encoding of stimulus strength using single electrical stimuli to the hindlimb, for which the strength could be reliably and precisely determined. Increasing stimulus strength under urethane anaesthesia resulted in a progressive increase in the population dendritic calcium response ($n = 21$; Fig. 1d; see Supplementary Fig. 1 for $\Delta F/F$ peak amplitude). Blockade of GABA receptors dramatically increased the dendritic population response ($n = 6$; Fig. 1d, e, red), possibly also because of runaway cortical excitation. Blockade of $GABA_B$ receptors alone had very little effect ($n = 6$; Fig. 1e, green). Normalization to the maximal dendritic response in each condition revealed that the responses to increasing stimulus strength were nearly linear under control conditions but highly nonlinear ('all or none') without $GABA_A$ receptor activity (Fig. 1f).

Conscious perception involves a combination of feed-forward and feedback processes[19] that may specifically influence dendritic encoding. To examine dendritic sensory-evoked responses in awake animals, we used a custom-built head mount to stabilize the fibre-optic cable[17]. As in the anaesthetized state, we found that increasing stimulus strength resulted in a graded increase in the population dendritic calcium response ($n = 3$; Fig. 1g). This shows that the mechanisms controlling the linear encoding of dendritic responses to sensory stimulus strength are not affected by anaesthesia. Thus, we continued our experiments to understand in detail what neural mechanisms underlie the change in linearity and dynamic range in anaesthetized animals.

We next investigated the source of dendritic Ca^{2+}. Application of D(-)-2-amino-5-phosphonovaleric acid (APV, to block NMDA receptors) to the cortical surface did not alter the amplitude of the dendritic population signal, indicating negligible contribution by subthreshold Ca^{2+} entry (that is, excitatory postsynaptic potentials (EPSPs)) as shown previously[17] ($P > 0.1$, $n = 6$; Supplementary Fig. 2). However, actively propagated action potentials from the soma into the apical dendrite can cause moderate increases in intracellular calcium concentration ($[Ca^{2+}]_i$), and dendritic Ca^{2+} spikes cause a large signal that is easily detectable[17]. There was no way to directly estimate action potential firing with the periscope method, so we tested the contribution of back-propagating action potentials (BPAPs) to the sensory-evoked dendritic calcium signals by injecting TTX (a Na^+ channel blocker) into L5 (Fig. 2a). Surprisingly, blocking activity in L5 and preventing BPAPs did not decrease the dendritic signal but rather increased it threefold ($1.2 \pm 0.61\%$ versus

[1]Physiologisches Institut, Universität Bern, Bühlplatz 5, CH-3012 Bern, Switzerland.
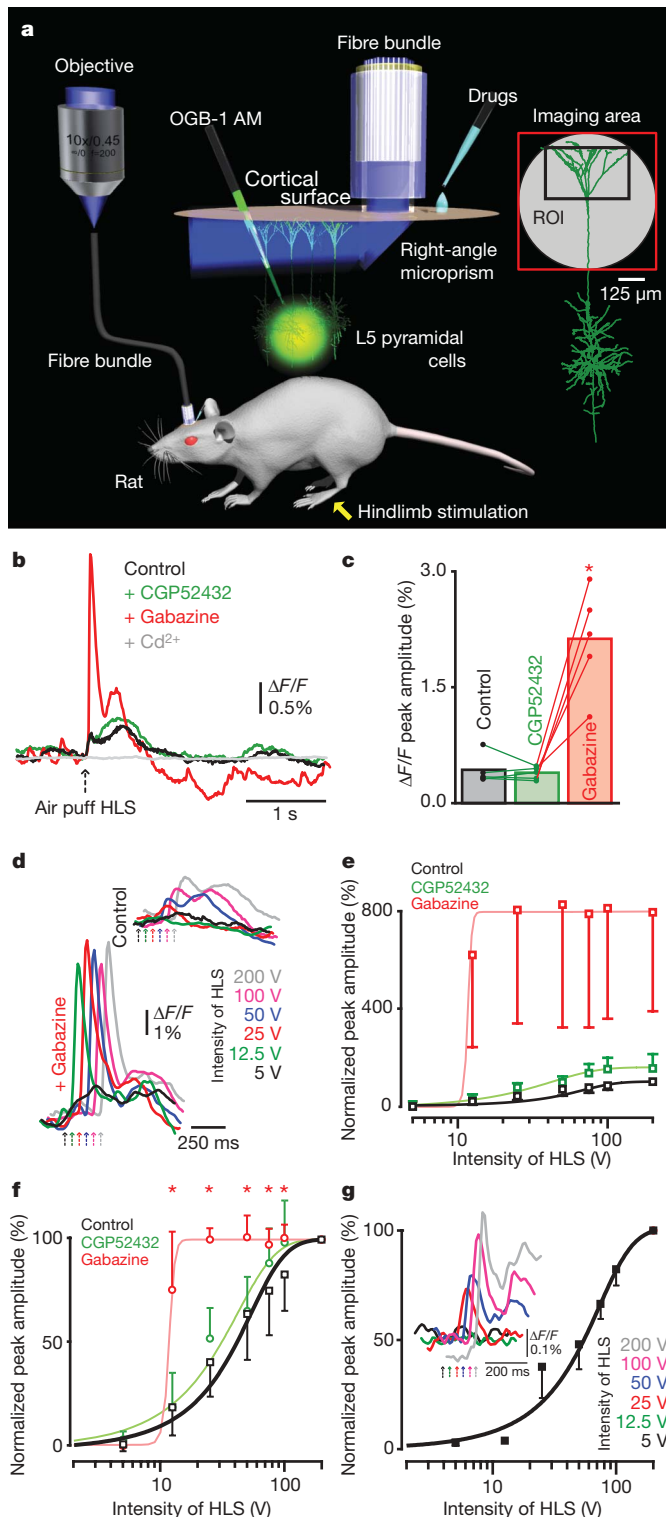
Figure 1 | Graded dendritic population Ca²⁺ responses to somatosensory inputs in anaesthetized and awake rats. a, Sketch of the experimental design using the periscope for recording specifically from the apical tuft dendrites of L5 neocortical pyramidal neurons[17]. Right: reconstructed pyramidal cell from *in vitro* experiments, demonstrating the typical regions of the dendritic tree within the imaging area. **b**, Averaged dendritic Ca²⁺ population signals (relative fluorescence change, $\Delta F/F$; 15 trials) recorded with the periscope following contralateral hindlimb stimulation with an air puff in anaesthetized rats (1 μM CGP52432, 3 μM gabazine and 1 mM Cd²⁺ applied to cortical surface). **c**, Summary of **b** ($n = 5$). Circles show the experiments and bars indicate their averages. **d**, Top: Averaged traces (15 trials) showing increase of dendritic Ca²⁺ signals following increase in contralateral electrical stimulation (arrows indicate the stimulus timing, time-shifted for clarity) of the hindlimb with increasing intensities (5–200 V) under control conditions. Bottom: same, after blockade of all GABA receptors. **e**, Summary of **d** ($n = 21$ for control, $n = 6$ for CGP52432 and gabazine; unpaired *t*-tests) fitted with sigmoidal curves with normalized data to maximal control value (200 V). **f**, Summary of **d** with data normalized to maximum in each condition (200 V). **g**, Summary of awake experiments ($n = 3$). Inset, averaged traces of dendritic Ca²⁺ signals in an awake animal. HLS, hindlimb stimulation; ROI, region of interest. *$P < 0.05$. Error bars, s.d.

receptors and kainate receptors) to L5 (see also Fig. 3). Again, the combined dendritic signal increased ($0.61 \pm 0.19\%$ versus $0.99 \pm 0.36\%$, $P < 0.05$, $n = 5$; Fig. 2c, f). Even more importantly, blockade of both L5 activity using TTX and disynaptic connectivity using CNQX had exactly the same effect on the slope of the dendritic response curve as a function of stimulus strength (Fig. 2h; see Supplementary Fig. 3 for $\Delta F/F$ peak amplitude). We conclude that deep cortical interneurons control dendritic encoding of sensory inputs.

This conclusion was further supported by the counterintuitive action of gabazine when applied locally to L5, which decreased the dendritic Ca²⁺ signal by nearly 40% ($0.76 \pm 0.46\%$ versus $0.45 \pm 0.30\%$, $P < 0.05$, $n = 5$; Fig. 2d, g, h). This is consistent with the hypothesis that disinhibition of L5 cells leads to an increase in dendritic inhibition in the upper layers. Moreover, muscimol (a GABA_A receptor agonist) application to L5 increased the signals evoked by hindlimb stimulations ($0.59 \pm 0.36\%$ versus $0.78 \pm 0.46\%$, $P < 0.05$, $n = 4$; Fig. 2i), whereas muscimol application to layer 1 (L1) decreased dendritic Ca²⁺ signals ($0.89 \pm 0.47\%$ versus $0.63 \pm 0.39\%$, $P < 0.05$, $n = 4$; Fig. 2i). This experiment suggests that inhibition is mainly dendritic, not perisomatic, which was confirmed with muscimol *in vitro* ($n = 3$, Supplementary Fig. 4). Last, these experiments demonstrate that this form of cortical inhibition predominates over increased firing in pyramidal neurons and increases in recurrent activity.

We next investigated the mechanisms underlying inhibitory control of dendritic activity. For this, we examined whether disynaptically evoked inhibition is sufficient to block dendritic Ca²⁺ spikes. We performed triple-patch recordings *in vitro* from pairs of disynaptically coupled L5 pyramidal neurons[20] (Fig. 3a) with the third recording in the dendritic Ca²⁺ spike initiation zone 500–800 μm from the cell body ($n = 3$, Fig. 3a). Trains of 18–20 action potentials at 80–100 Hz in the presynaptic pyramidal neuron evoked disynaptic dendritic inhibition (Fig. 3b). Dendritic Ca²⁺ spikes evoked with dendritic current injection (Fig. 3c) could be completely abolished by the disynaptically evoked dendritic inhibition in an all-or-none fashion (Fig. 3d). The same phenomenon could be observed with calcium imaging of the Ca²⁺ spike initiation zone. Disynaptically evoked inhibition severely reduced the Ca²⁺ fluorescence due to Ca²⁺ spikes into this region evoked with high-frequency trains of back-propagating action potentials[22] ($23.0 \pm 19.8\%$, $P < 0.05$, $n = 3$; Fig. 3e–g; for details, see Supplementary Fig. 5). The effect of disynaptic inhibition was blocked by local application of CNQX to L5 analogous to the *in vivo* experiments ($n = 3$). The strongest blockade of inhibition occurred when CNQX was puffed close to L5 ($n = 3$, Fig. 3h–j), implying that most of the disynaptic dendritic inhibition is evoked by cells in lower cortical layers.

$3.3 \pm 0.95\%$, $P < 0.05$, $n = 5$; Fig. 2b, e) and changed the biphasic nature of the dendritic response. This result shows first that the large calcium responses detected are dependent on neither dendritic BPAPs nor axonal action potentials and second that an additional TTX-sensitive mechanism in L5 suppresses dendritic [Ca²⁺]_i.

The most likely explanation is that TTX also blocked some inhibitory neurons in L5. Recent studies *in vitro* have shown that inhibition of pyramidal dendrites can be evoked by activation of dendrite-targeting Martinotti interneurons by neighbouring pyramidal neurons[20,21]. To test this possibility, we suppressed the activation of this form of disynaptic inhibition by local injection of CNQX (an antagonist of α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)
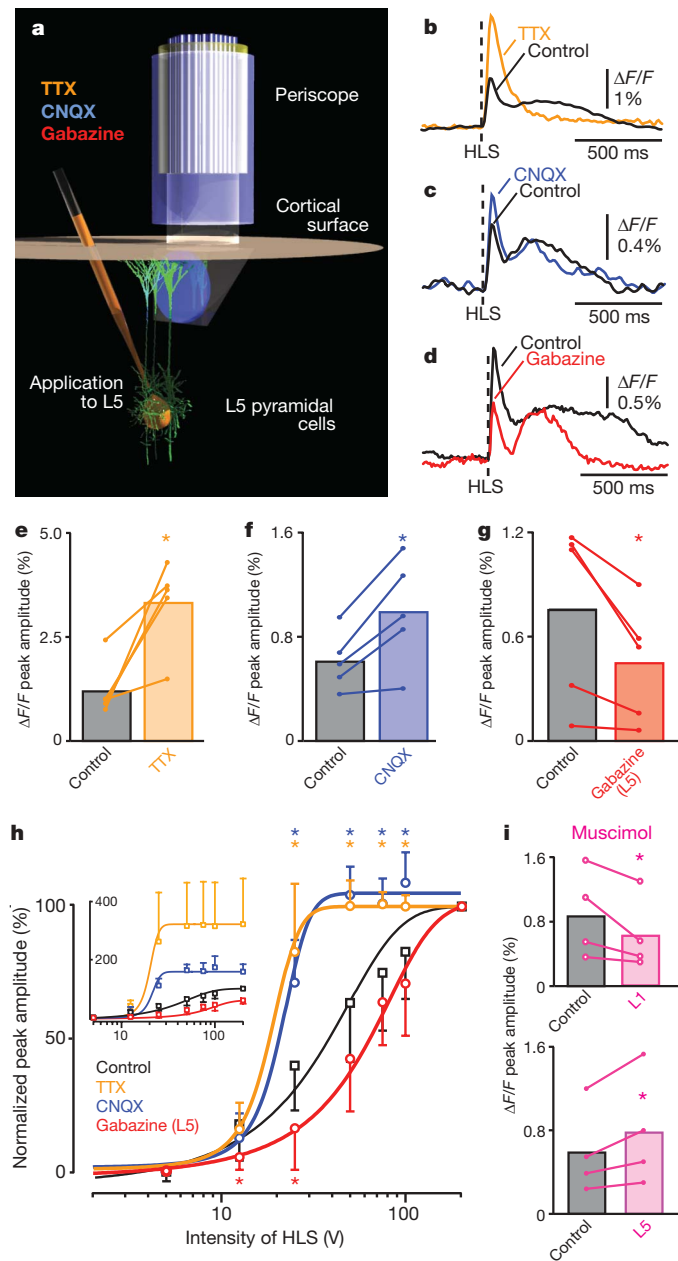
1138

Figure 2 | Deep-layer control of dendritic activity. a, Sketch showing local application of tetrodotoxin (TTX, 3 μM), 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX, 100 μM) and gabazine (1 μM) injected into L5. **b–d** Averaged responses (15 trials) in single animals. **e–g,** Summary of **b–d** (n = 5 for each drug). **h,** Responses (normalized to maximum of each condition) versus stimulus strength in control (n = 21), with TTX (n = 5), CNQX (n = 5) and gabazine (n = 5). Asterisks indicate significance with unpaired t-tests. Inset, data normalized to control. **i,** Summary of muscimol injection to L1 (top, n = 4) and to L5 (bottom, n = 4). Application of normal rat ringer to L5 and L1 did not change evoked Ca^{2+} signals (0.83 ± 0.76% for control versus 0.76 ± 0.49% for L5 and 0.77 ± 0.46% for L1, n = 3). *P < 0.05. Error bars, s.d.

The experiments show that very specific cortical microcircuitry is involved in determining the responsiveness of apical dendrites of L5 pyramidal neurons (Fig. 4a). Both the *in vivo* and the *in vitro* data suggest that individual dendritic Ca^{2+} signals contribute to the population responses by means of all-or-none Ca^{2+} spikes consistent with previous recordings *in vivo*[8,23], and that the inhibition of dendritic Ca^{2+} spikes is also all or none[24,25]. To investigate how the microcircuitry shapes the grading of the dendritic population response, we modelled a population of two-compartment pyramidal neurons[5]
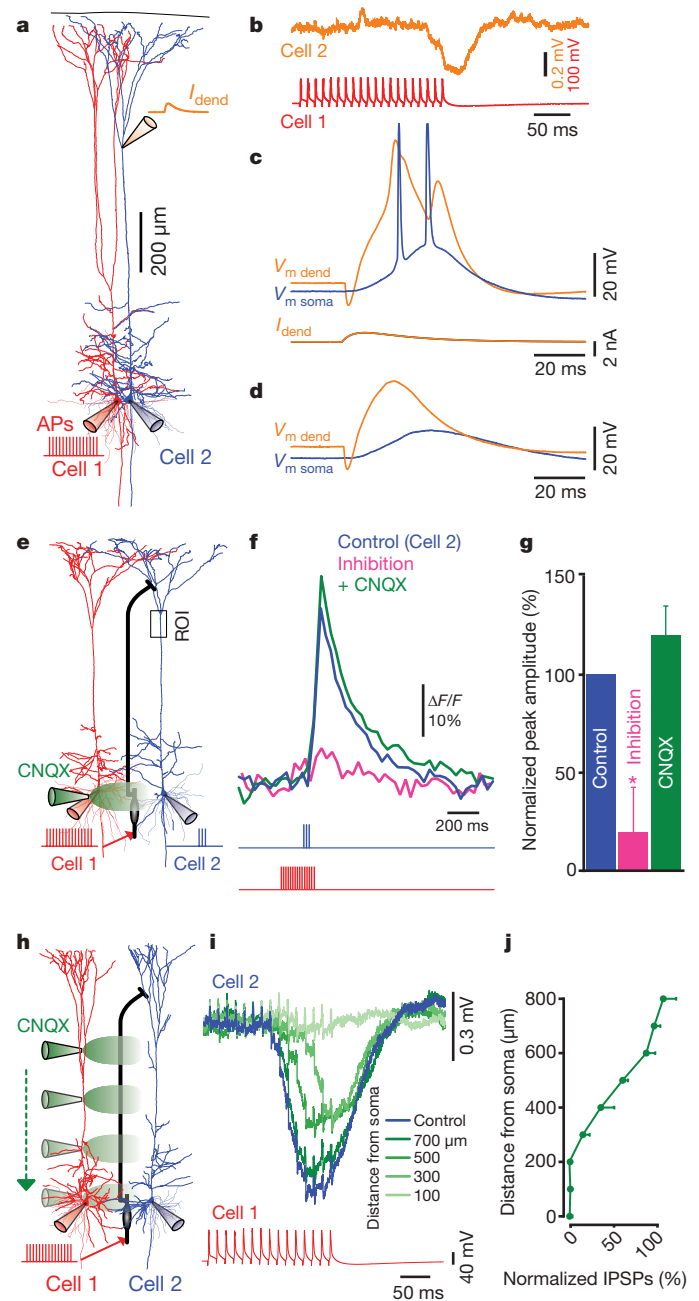
Figure 3 | Disynaptic inhibition blocks dendritic Ca^{2+} spikes *in vitro*. a, Experimental diagram showing reconstructed L5 somatosensory pyramidal neurons and recording sites. **b,** Averaged disynaptically evoked inhibitory postsynaptic potential (IPSP) (15 sweeps). **c,** Dendritic Ca^{2+} spike ($V_{m\ dend}$) evoked by dendritic current injection (I_{dend}) caused a burst of somatic action potentials ($V_{m\ soma}$). **d,** Same as **c,** but with disynaptic inhibition. **e,** Experimental diagram. Pyramidal neurons were artificially separated and the putative interneuron (black) added schematically. Cell 2 was filled with OGB-1 (100 μM) and Ca^{2+} fluorescence ($\Delta F/F$) measured at a distal ROI. CNQX (20 μM) was applied 100 μm from the soma. **f,** Ca^{2+} fluorescence transients (blue) were blocked by disynaptic inhibition (pink) and recovered with local application of CNQX (green). **g,** Summary of **f** (n = 3). **h,** Experimental diagram. CNQX was applied at nine locations 100 μm apart in the vertical axis. **i,** Averaged IPSP (30 sweeps) recorded in cell 2 evoked by a train of presynaptic action potentials in cell 1. **j,** Summary of **i** (n = 3). AP, action potential. *P < 0.05. Error bars, s.d.

receiving distributed feed-forward and feedback excitation and inhibition (Fig. 4b; see Methods and Supplementary Information, model description). The distribution of Ca^{2+} spike thresholds in the dendritic compartments (Fig. 4c, d, green) was chosen to allow for a steep all-or-none Ca^{2+} population response when dendritic
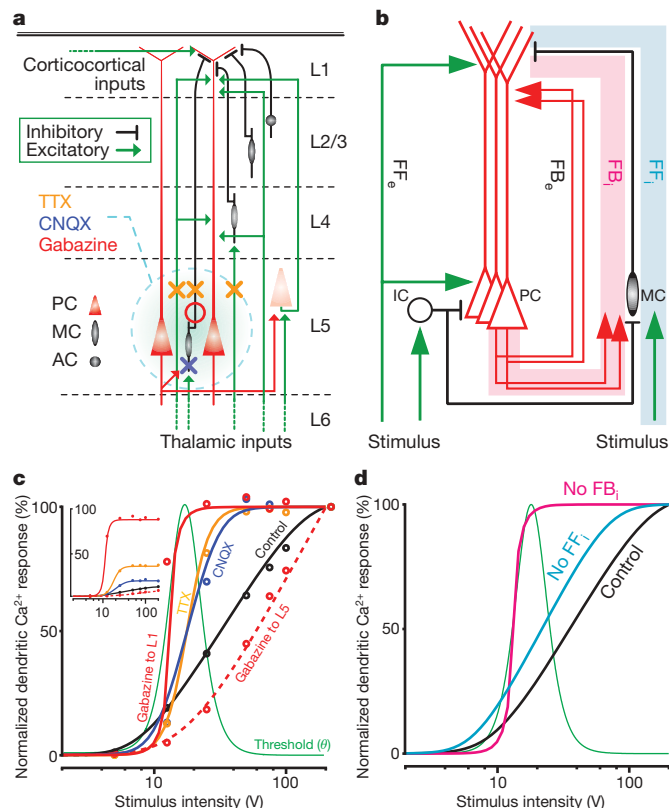
Figure 4 | Model of microcircuitry with all-or-none dendritic Ca²⁺ spikes.
a, Diagram of cortical circuitry. The blue shaded region indicates the
elements in the circuit affected by TTX, CNQX and gabazine. PC, pyramidal
cell; MC, Martinotti cell; AC, aspiny cell. **b**, Model microcircuitry with a
population of two-compartment pyramidal neurons with all-or-none
dendritic Ca²⁺ events and four major classes of dendritic inputs (FF$_e$, feed-
forward excitation; FF$_i$, feed-forward inhibition; FB$_e$, feedback excitation;
FB$_i$, feedback inhibition; Blue and green arrows, excitation; black lines,
inhibition). Dendritic feed-forward inhibition and feedback inhibition both
act through deep-layer Martinotti cells (red and blue stripes). IC, inhibitory
cell. **c**, Data fitted with the model by mimicking the pharmacological
blocking experiments in L5 and L1 (see Supplementary Information, model
description). Inset, model responses scaled to the number of neurons in the
population recruitable for Ca²⁺ spikes. **d**, Effects of blocking feedback and
feed-forward inhibition, respectively (see text and also Supplementary Fig.
10, which shows the effect of changing the strength of inhibition on the
slope).

inhibition is blocked (for example for gabazine to the cortical surface;
Fig. 4c). The dendritic population response curves for the pharmaco-
logical blocking experiments could be reproduced by removing the
corresponding model connectivity to upper and lower layers,
respectively (Fig. 4b, c). The model also predicts that blocking the
feed-forward drive onto the inhibitory Martinotti cells[26] merely shifts
the response curve to the left by reducing its threshold (Fig. 4d, light
blue), whereas blocking feedback inhibition[20,21] increases the gain of
the response curve and thus modulates the dynamic range of the
population calcium response (Fig. 4d, pink; see Supplementary Fig.
11 for further details).

We conclude that the representation of sensory stimuli by cortical
output neurons cannot be fully explained by the traditional axoso-
matic integration approach, but requires active dendrites embedded
within highly specialized cortical microcircuitry. Dendritic activity is
tightly controlled by interneurons projecting from deep to upper
layers (Martinotti cells) whose function was previously unknown.
We show here that these neurons dynamically modulate the slope
and threshold of the dendritic response function to match the
physiologically relevant input range.

METHODS SUMMARY

Female Wistar rats (24–40 d old) were used for these experiments. Multiple
dendrites of L5 pyramidal cells were loaded with a Ca²⁺-sensitive dye, Oregon
Green 488 BAPTA-1 AM (OGB-1 AM), by means of bolus injection to L5 (ref.
17). Fibre-optic calcium imaging *in vivo* was performed in the primary somato-
sensory cortex of awake and urethane-anaesthetized rats as described prev-
iously[17] (Fig. 1). Briefly, we used a custom-built microscope attached to a
fibre-optic cable (IGN-06/17, 680-μm diameter, Sumitomo Electric
Industries). A GRIN lens, right-angled prism (500 × 500 μm; GrinTech) assem-
bly was attached to the end of the fibre and the prism inserted into the hindlimb
area of the somatosensory cortex (determined beforehand with intrinsic
imaging). Sensory responses were evoked by brief air puffs (50-ms duration)
delivered to the contralateral hindlimb or a single short electrical microstimula-
tion pulse (0.1-ms duration, 5–200 V) through a surface electrode. All stimuli
were below the threshold for muscle responses in the hindlimb. *In vitro* patch-
clamp recordings from L5 pyramidal cells were made in parasagittal slices pre-
pared with standard techniques[25]. Statistical analysis was performed with paired
t-tests unless otherwise noted; asterisks denote $P < 0.05$.

Full Methods and any associated references are available in the online version of
the paper at www.nature.com/nature.

1. London, M. & Häusser, M. Dendritic computation. *Annu. Rev. Neurosci.* **28,**
    503–532 (2005).
2. Llinás, R. R. The intrinsic electrophysiological properties of mammalian neurons:
    insights into central nervous system function. *Science* **242,** 1654–1664 (1988).
3. Johnston, D., Magee, J. C., Colbert, C. M. & Christie, B. R. Active properties of
    neuronal dendrites. *Annu. Rev. Neurosci.* **19,** 165–186 (1996).
4. Destexhe, A., Mainen, Z. F. & Sejnowski, T. J. Synthesis of models for excitable
    membranes, synaptic transmission and neuromodulation using a common kinetic
    formalism. *J. Comput. Neurosci.* **1,** 195–230 (1994).
5. Larkum, M. E., Senn, W. & Lüscher, H.-R. Top-down dendritic input increases the
    gain of layer 5 pyramidal neurons. *Cereb. Cortex* **14,** 1059–1070 (2004).
6. Schiller, J., Schiller, Y., Stuart, G. & Sakmann, B. Calcium action potentials
    restricted to distal apical dendrites of rat neocortical pyramidal neurons. *J. Physiol.
    (Lond.)* **505,** 605–616 (1997).
7. Kim, H. G. & Connors, B. W. Apical dendrites of the neocortex: correlation
    between sodium- and calcium-dependent spiking and pyramidal cell morphology.
    J. Neurosci. **13,** 5301–5311 (1993).
8. Larkum, M. E. & Zhu, J. J. Signaling of layer 1 and whisker-evoked Ca2+ and Na+
    action potentials in distal and terminal dendrites of rat neocortical pyramidal
    neurons *in vitro* and *in vivo. J. Neurosci.* **22,** 6991–7005 (2002).
9. Williams, S. R. & Stuart, G. J. Mechanisms and consequences of action potential
    burst firing in rat neocortical pyramidal neurons. *J. Physiol. (Lond.)* **521,** 467–482
    (1999).
10. White, E. L. & Hersch, S. M. A quantitative study of thalamocortical and other
    synapses involving the apical dendrites of corticothalamic projection cells in
    mouse SmI cortex. *J. Neurocytol.* **11,** 137–157 (1982).
11. Hersch, S. M. & White, E. L. Thalamocortical synapses with corticothalamic
    projection neurons in mouse SmI cortex: electron microscopic demonstration of a
    monosynaptic feedback loop. *Neurosci. Lett.* **24,** 207–210 (1981).
12. Zhu, Y. & Zhu, J. J. Rapid arrival and integration of ascending sensory information
    in layer 1 nonpyramidal neurons and tuft dendrites of layer 5 pyramidal neurons of
    the neocortex. *J. Neurosci.* **24,** 1272–1279 (2004).
13. Oda, S. *et al.* Thalamocortical projection from the ventral posteromedial nucleus
    sends its collaterals to layer I of the primary somatosensory cortex in rat. *Neurosci.
    Lett.* **367,** 394–398 (2004).
14. Budd, J. M. L. Extrastriate feedback to primary visual cortex in primates: a
    quantitative analysis of connectivity. *Proc. R. Soc. Lond. B* **265,** 1037–1044 (1998).
15. Cauller, L. J. & Connors, B. W. Synaptic physiology of horizontal afferents to layer-I
    in slices of rat SI neocortex. *J. Neurosci.* **14,** 751–762 (1994).
16. Elhanany, E. & White, E. L. Intrinsic circuitry: synapses involving the local axon
    collaterals of corticocortical projection neurons in the mouse primary
    somatosensory cortex. *J. Comp. Neurol.* **291,** 43–54 (1990).
17. Murayama, M., Pérez-Garci, E., Lüscher, H. R. & Larkum, M. E. Fiberoptic system
    for recording dendritic calcium signals in layer 5 neocortical pyramidal cells in
    freely moving rats. *J. Neurophysiol.* **98,** 1791–1805 (2007).
18. Kerr, J. N., Greenberg, D. & Helmchen, F. Imaging input and output of neocortical
    networks *in vivo. Proc. Natl Acad. Sci. USA* **102,** 14063–14068 (2005).
19. Cauller, L. Layer I of primary sensory neocortex: Where top-down converges upon
    bottom-up. *Behav. Brain Res.* **71,** 163–170 (1995).
20. Silberberg, G. & Markram, H. Disynaptic inhibition between neocortical pyramidal
    cells mediated by Martinotti cells. *Neuron* **53,** 735–746 (2007).
21. Kapfer, C., Glickfeld, L. L., Atallah, B. V. & Scanziani, M. Supralinear increase of
    recurrent inhibition during sparse activity in the somatosensory cortex. *Nature
    Neurosci.* **10,** 743–753 (2007).

22. Larkum, M. E., Kaiser, K. M. & Sakmann, B. Calcium electrogenesis in distal apical dendrites of layer 5 pyramidal cells at a critical frequency of back-propagating action potentials. *Proc. Natl Acad. Sci. USA* **96,** 14600–14604 (1999).

23. Helmchen, F., Svoboda, K., Denk, W. & Tank, D. W. *In vivo* dendritic calcium dynamics in deep-layer cortical pyramidal neurons. *Nature Neurosci.* **2,** 989–996 (1999).

24. Larkum, M. E., Zhu, J. J. & Sakmann, B. A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature* **398,** 338–341 (1999).

25. Pérez-Garci, E., Gassmann, M., Bettler, B. & Larkum, M. E. The GABAB1b isoform mediates long-lasting inhibition of dendritic Ca2+ spikes in layer 5 somatosensory pyramidal neurons. *Neuron* **50,** 603–616 (2006).

26. Tan, Z., Hu, H., Huang, Z. J. & Agmon, A. Robust but delayed thalamocortical activation of dendritic-targeting inhibitory interneurons. *Proc. Natl Acad. Sci. USA* **105,** 2187–2192 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions M.M. and M.E.L. designed the study. M.M. performed the periscope experiments *in vivo*, E.P.-G. performed the *in vitro* experiments, and T.N. and M.M. performed the *in vivo* two-photon experiments. W.S. and T.B. made the model and the supplementary model description. M.M. and M.E.L. prepared the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.E.L. (larkum@pyl.unibe.ch).

METHODS

The following methods have already been mostly described in ref. 17.

Animals and surgery. Female Wistar rats (P24–P40) were used in these experiments. Urethane (intraperitoneal, $1.5\,\mathrm{g\,kg}^{-1}$) was used for experiments under anaesthesia. The head was fixed in a stereotaxic instrument (Model SR-5R, Narishige) and body temperature maintained at 36 to 37 °C. A craniotomy above the primary somatosensory cortex (3×4.4 mm square), centred at 1.5 mm posterior to bregma and 2.2 mm from midline in the right hemisphere, was performed and the dura mater surgically removed immediately before Ca^{2+} recording (see below). For awake experiments, the scalp was removed under general anaesthesia (isoflurane, Baxter) and a local anaesthetic (lidocaine; Sigma-Aldrich) was applied to the wound. Following surgery, an analgesic was administered (buprenorphine (twice per day), Essex Chemie) and local anaesthetic applied to the scalp. On the day of the experiment, a craniotomy was performed, a Ca^{2+}-sensitive dye was injected into L5 (see below) and a metal post was fixed to the skull with dental cement. After anaesthesia, the animals were retrained in a holder (Clam005, Kent Scientific).

In vivo loading of Ca^{2+}-sensitive dye. Oregon Green 488 BAPTA-1 (OGB-1) AM (50 µg; Molecular Probes) was mixed with 5 µl of pluronic acid (Pluronic F-127, 20% solution in dimethylsulphoxide; Molecular Probes) for 15 min. The solution was then diluted in 28 µl of HEPES-buffered solution (125 mM NaCl, 2.5 mM KCl, 10 mM HEPES) and mixed for a further 15 min. The OGB-1 AM solution was loaded into a glass pipette (tip diameter, 5–20 µm) and pressure-injected into L5 (pressure, 10–25 kPa) for 1 min. The pipette was withdrawn and the area of the craniotomy was then submerged with rat ringer (135 mM NaCl, 5.4 mM KCl, 1.8 mM CaCl2, 1 mM MgCl2, 5 mM HEPES) for 2 h.

In vivo Ca^{2+} recording (periscope). A 100-W mercury lamp (U-LH100HG, Olympus) or a blue light-emitting-diode (LED, IBF+LS30W-3W-Slim-RX, Imac) was used as a light source. An excitation filter, a dichroic mirror, and an emission filter (as a filter set 31001, Chroma Technology) were used for epifluorescence Ca^{2+} recordings. A ×10 objective (0.45 numerical aperture, Model E58-372, Edmund Optics) was used for illuminating and imaging a fibre bundle (see below). A cooled charge-coupled-device (CCD) camera (MicroMax, Roper Scientific) was used for collecting fluorescence.

A fibre bundle (IGN-06/17, Sumitomo Electric Industries) consisting of 17,000 fibre elements was used as a combined illuminating–recording fibre. The end face of the bundle was fitted with a prism-lens assembly that consisted of a right-angled prism ($0.5 \times 0.5 \times 0.5$ mm; GrinTech) attached to a GRIN lens (diameter, 0.5 mm; 0.5 numerical aperture; GrinTech). The working distance was nominally 100 µm and the magnification was ×0.73, resulting in a field of view of 685-µm diameter.

Sensory responses were evoked by a brief air puff (50-ms duration) delivered to the contralateral hindlimb or a single short electrical stimulation (0.1-ms duration, 5–200 V). Fluorescence changes were sampled at 100 Hz. Data were acquired on a personal computer using WinView software (Roper Scientific). Regions of interest were chosen offline for measuring fluorescence changes (see data analysis section). Gabazine and CGP52432 were bought from Biotrend Chemicals, tetrodotoxin (TTX) from Tocris Cookson and cadmium chloride from Fluka.

In vivo Ca^{2+} recording (two-photon). Two-photon excitation fluorescence microscopy was either performed with a femtosecond infrared laser ($\lambda = 800$–810 nm, 100 fs; Tsunami pumped by a Millennia VIII or Mai-Tai HP, Spectra-Physics) coupled to a laser scanning microscope (TCS SP2RS, Leica) equipped with a ×40 water immersion objective (HCX APO w40x, UVI, numerical aperture 0.8) or a custom-built two-photon microscope equipped with a ×20 water immersion objective (numerical aperture 0.95; Olympus). Excitation infrared-laser light and fluorescence-emission light were separated at 670 nm (excitation filter 670DCXXR, AHF Analysentechnik) and the infrared light was blocked in the detection pathway with an infrared-block filter (E700SP, AHF Analysentechnik). Fluorescence was detected using epifluorescence non-descanned detection. Calcium transients were recorded in line-scan mode at 500 Hz from identified L5 tuft dendrites.

In vitro experiments. Animals were anaesthetized with a mixture of 95% CO_2/5% O_2 before decapitation. The brain was then rapidly transferred to ice-cold, oxygenated artificial cerebrospinal fluid (ACSF) containing 125 mM NaCl, 25 mM NaHCO3, 2.5 mM KCl, 1.25 mM NaH2PO4, 1 mM MgCl2, 25 mM glucose and 2 mM CaCl2 (pH 7.4). Parasagittal slices of the primary somatosensory cortex (300-µm thick) were cut with a vibrating microslicer on a block angled at 15° to the horizontal and maintained at 37 °C in the preceding solution for 15–120 min before use.

Ca^{2+} imaging and somatic whole-cell patch recordings of L5 neurons were obtained with Nikon Eclipse E600FN microscopes (Nikon). Ca^{2+} imaging was performed using a CCD camera (CoolSNAP, Roper Scientific). The fluorescence was observed using standard epifluorescence filter sets for fluorescein (excitation at 480 nm; used for OGB-1) and Texas Red (Alexa 594; Molecular Probes) fluorescence (Chroma Technology). Fluorescence intensities were sampled at 30–40 Hz. Pipettes (4–6 MΩ) for the whole-cell patch-clamp recordings were filled with an intracellular solution containing 135 mM K-gluconate, 7 mM KCl, 10 mM HEPES, 10 mM Na2-phosphocreatine, 4 mM Mg-ATP, 0.3 mM GTP, 10 mM Alexa 594 and 0.2% biocytin (pH 7.2 with KOH). No correction was made for the junction potential between the bath and pipette solutions. The recordings were made with Multiclamp 700B (Axon Instruments), digitized at 10 kHz with an analogue–digital converter (ITC-16 Instrutech) and acquired on a personal computer using Igor software (WaveMetrics). Dendritic Ca^{2+} spikes were evoked with EPSP-shaped current waveforms injected through a dendritic pipette: double exponential, $f_{(t)} = (1 - e^{-t/\tau_1})e^{-t/\tau_2}$, where $\tau_1 = 4$ ms and $\tau_2 = 10$ ms; time to peak, 5 ms. Slices were perfused continuously with the ACSF at 33–35 °C throughout the experiments.

Data analysis. The fluorescence signals *in vivo* were quantified by measuring the mean pixel value of a manually selected (offline) ROI for each frame of the image stack using Igor software. Ca^{2+} changes were expressed as $\Delta F/F = F_t/F_0$, where F_t was the average fluorescence intensity within the ROI at time t during the imaging experiment and F_0 was the mean value of fluorescence intensity before stimulation. For *in vitro* experiments, Ca^{2+} changes were expressed as $\Delta F/F = (F_t - F_0)/(F_0 - F_B)$, where F_B is the background fluorescence measured from a region away from the recorded area.

Model. The full model comprises microcircuitry with two types of inhibitory neuron and a population of N two-compartment pyramidal neurons displaying all-or-none dendritic calcium events (0 or c). The mathematical analysis of the microcircuitry leads to a compact description of the population calcium response, $C(S)$, as a function of the stimulus strength, S, and the feed-forward and feedback connection strengths $\beta^{ff} = FF_e - FF_i$ and $\beta^{fb} = FB_e - FB_i$, respectively. We fit the distribution of the calcium activation thresholds (mean, θ_0; standard deviation, σ) such that the steep response curve for TTX (modelled by $\beta^{fb} = 0$) is reproduced. The population calcium response can be approximated by the sigmoidal function

$$C(S) \approx \frac{Nc}{1 + \exp\left(-\gamma(\beta^{ff}\log S - (\theta_0 + \theta_1))/\sigma\right)}$$

with gain factor $\gamma = (1 - \beta^{fb}c/4\sigma)^{-1}$ and threshold shift $\theta_1 = -\beta^{fb}c/2$; see Supplementary Information, model description. The pharmacological blocking experiments are described by assuming that the corresponding model connections are removed (Fig. 4b, c), and this changes the values of β^{ff} and β^{fb}. Roughly speaking, these experiments lead to an increase in β^{fb} (CNQX, TTX, gabazine to L1) with a corresponding gain increase, and to a reduction in β^{ff} (gabazine to L5) with a corresponding rightwards shift of the $C(S)$ curve. Analogously, blocking the feed-forward excitation of the (inhibitory) model Martinotti cells increases β^{ff} and leads to a leftwards shift of $C(S)$, whereas blocking the Martinotti cell output increases β^{fb} and leads to a gain increase (blue and pink curves in Fig. 4d). The number of pyramidal neurons, N, that can be recruited to contribute to the population calcium signal is a nonlinearly increasing function of the dendritic feedback strength β^{fb} and a linearly increasing function of the gain factor γ (see Supplementary Information, model description).

LETTERS

The subcellular organization of neocortical excitatory connections

Leopoldo Petreanu[1], Tianyi Mao[1], Scott M. Sternson[1] & Karel Svoboda[1]

Understanding cortical circuits will require mapping the connections between specific populations of neurons[1], as well as determining the dendritic locations where the synapses occur[2]. The dendrites of individual cortical neurons overlap with numerous types of local and long-range excitatory axons, but axodendritic overlap is not always a good predictor of actual connection strength[3–5]. Here we developed an efficient channelrhodopsin-2 (ChR2)-assisted method[6–8] to map the spatial distribution of synaptic inputs, defined by presynaptic ChR2 expression, within the dendritic arborizations of recorded neurons. We expressed ChR2 in two thalamic nuclei, the whisker motor cortex and local excitatory neurons and mapped their synapses with pyramidal neurons in layers 3, 5A and 5B (L3, L5A and L5B) in the mouse barrel cortex. Within the dendritic arborizations of L3 cells, individual inputs impinged onto distinct single domains. These domains were arrayed in an orderly, monotonic pattern along the apical axis: axons from more central origins targeted progressively higher regions of the apical dendrites. In L5 arborizations, different inputs targeted separate basal and apical domains. Input to L3 and L5 dendrites in L1 was related to whisker movement and position, suggesting that these signals have a role in controlling the gain of their target neurons[9]. Our experiments reveal high specificity in the subcellular organization of excitatory circuits.

We recorded from pyramidal neurons in neocortical brain slices containing ChR2- expressing axons[7] (Fig. 1a). To map the dendritic locations of input from ChR2-positive axons (Supplementary Fig. 1), we used a laser to depolarize these axons only in the vicinity of the laser beam (that is, with action potentials blocked), triggering local glutamate release (subcellular ChR2-assisted circuit mapping (sCRACM)). We blocked Na$^+$ channels (1 μM tetrodotoxin (TTX)) and the K$^+$ channels that are critical for repolarization of the axon (200 nM α-dendrotoxin or 100 μM 4-aminopyridine (4-AP)) (Fig. 1b)[10]. Under these conditions, photostimulation with short (1-ms) light pulses (<2 mW) triggered robust excitatory postsynaptic currents (EPSC$_{sCRACM}$; Fig. 1b). Higher light intensities caused larger EPSC$_{sCRACM}$ amplitudes (Fig. 1c) and shorter onsets (Supplementary Fig. 2). As the cylindrical laser beam was scanned over the dendrites of a recorded neuron (map pattern: 12 × 24 grid, 50-μm spacing), EPSC$_{sCRACM}$ were detected only when the laser beam overlapped with the dendritic arborization of the recorded cell and with ChR2-positive axons (Fig. 1b, d–f), indicating that under these conditions light depolarizes ChR2-positive axons to cause local release of neurotransmitter. Converting EPSC$_{sCRACM}$ into pixel values
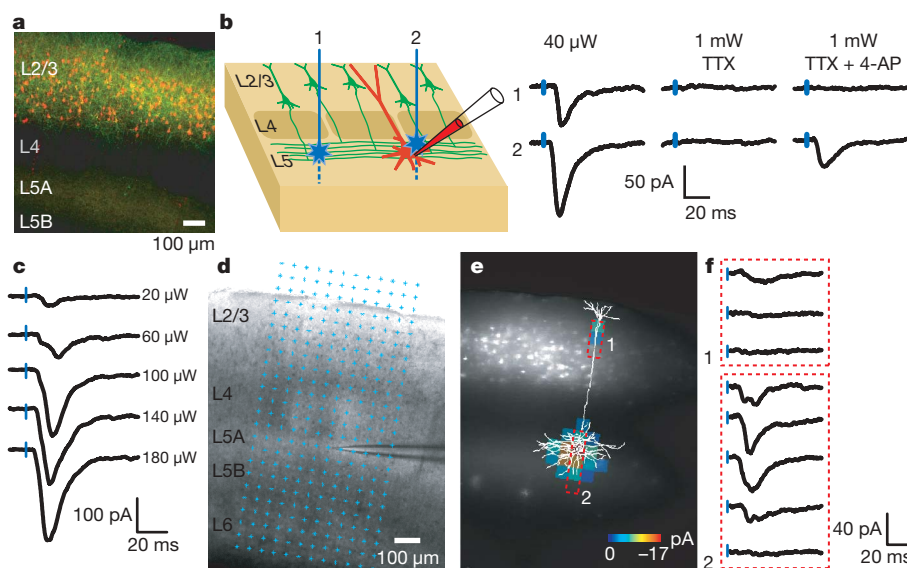
Figure 1 | Subcellular ChR2-assisted circuit mapping. a, Confocal image showing L2/3 neurons expressing mCherry (red) and ChR2-Venus (green) in the barrel cortex. b, Left: sketch of the photostimulation geometry. Right: excitatory postsynaptic currents evoked by photostimuli corresponding to the locations indicated in the sketch. Blue tickmarks indicate the laser pulse. Laser power is indicated at top. c, EPSC$_{sCRACM}$ evoked by photostimulation with increasing laser powers (right). d, Bright-field image of a brain slice showing the recording pipette and the photostimulation grid (blue dots). e, sCRACM map overlaid on a fluorescence image, showing ChR2-positive neurons, and the reconstructed dendrite of the recorded neuron (same experiment as d). Non-zero EPSC$_{sCRACM}$ are colour-coded to represent mean amplitude. f, EPSC$_{sCRACM}$ recorded in the boxed regions in e.

(EPSC$_{sCRACM}$ averaged over a time window 0–75 ms post-stimulus) provides a two-dimensional 'image' of the distribution of specific input within the dendritic arborizations of the recorded cell (Fig. 1e, f). Measurements of the point spread function revealed that sCRACM maps specific types of input with ~60-μm spatial resolution (Supplementary Fig. 3).

EPSC$_{sCRACM}$ amplitudes depend on the density of ChR2-positive axons, the fraction of axons that make synapses with the recorded neuron, the strength of the synapses and their electrotonic distances from the soma[11] (see Methods). Because the density of ChR2-positive axons varies between preparations, sCRACM maps were normalized to the largest pixels within a map and thus represent the relative strengths of input within the dendritic tree. Repeated sCRACM maps were reproducible at the level of single pixels (Supplementary Fig. 4), and the structure of peak-normalized maps was similar across a large range of light intensities (Supplementary Fig. 4).

Multiple types of axon overlap with the dendrites of cortical neurons[5,12–17]. Some axons arise locally[5,14,18], whereas others ascend from the thalamus[15,16] or descend from higher cortical areas[17]. To determine which axons in a target region connect with a particular cell type, and the spatial distributions of specific inputs within the dendritic tree, we expressed ChR2 in five distinct axonal populations (in separate experiments) that overlap with pyramidal cell dendrites in the barrel cortex (Supplementary Fig. 5).

The spatial distribution of labelled axons in the barrel cortex was largely in agreement with previous anatomical studies (Supplementary Fig. 5). Projections from the ventral posterior medial nucleus (VPM) of the thalamus were focused in L4 and at the border of L5 with L6 (Fig. 2a, Supplementary Fig. 5)[16], but diffuse axons were found throughout all cortical layers[15]. L4 axons arborized in L4 and ascended into L2/3, up to the lower edge of L1 (ref. 18); a weaker projection descended into L5 and L6 (ref. 18). Axons from L2/3 pyramidal cells arborized within L2/3 and on the border of L5A with

L5B (refs 7, 14). Axons from the primary whisker motor cortex (M1) arborized densely in L1, and more diffusely in L5 and L6 (ref. 17). A dense bundle of ascending M1 axons was often apparent next to the most medial barrels (Fig. 2a, arrowhead). Axons from the posterior medial nucleus (POm) of the thalamus were focused in L5A, and more weakly in L1 (ref. 15).

We mapped specific types of input within the dendritic trees of individual L3 cells (Fig. 2a). Maps were then averaged across cells either aligned on the pia (Fig. 2b), to visualize the laminar location of the inputs, or aligned on the soma, to measure the location of the inputs relative to the soma (Fig. 2c). L3 cells received input from all five projections. Each input overlapped with a single contiguous dendritic subregion (Fig. 2c, d). Ascending input from VPM to L3 (VPM → L3) was focused on the bottom part of the basal dendritic arborizations. Input from ascending L4 → L3 axons was centred on the soma and basal arborization, above the input from VPM. Input from recurrent L2/3 → L3 axons was mostly in the upper basal dendrites and the apical oblique dendrites, above the input from L4 (see also Supplementary Fig. 6). Feedback from M1 targeted the tuft branches in L1, above the input from L2/3. The positions of VPM, L4, L2/3 and M1 input along the apical axis of L3 neurons mirrors the flow of excitation within the cortical circuit: more peripheral input impinges on lower parts of the dendritic arborization, whereas more central input impinges on higher parts of the dendritic arborization. During somatosensation, L3 neurons thus receive an ascending wave of excitatory input. POm → L3 input was weighted towards L1, although it was distributed relatively broadly, spanning most of the dendritic arborization.

We next mapped the same group of five inputs within the dendrites of L5 pyramidal neurons. Both L5A and L5B pyramidal cells received input from L4, L2/3, M1 and VPM (Fig. 3a, b, Supplementary Fig. 7). To quantify the strength of input from defined axonal projections across postsynaptic cells in different layers, we recorded from pairs
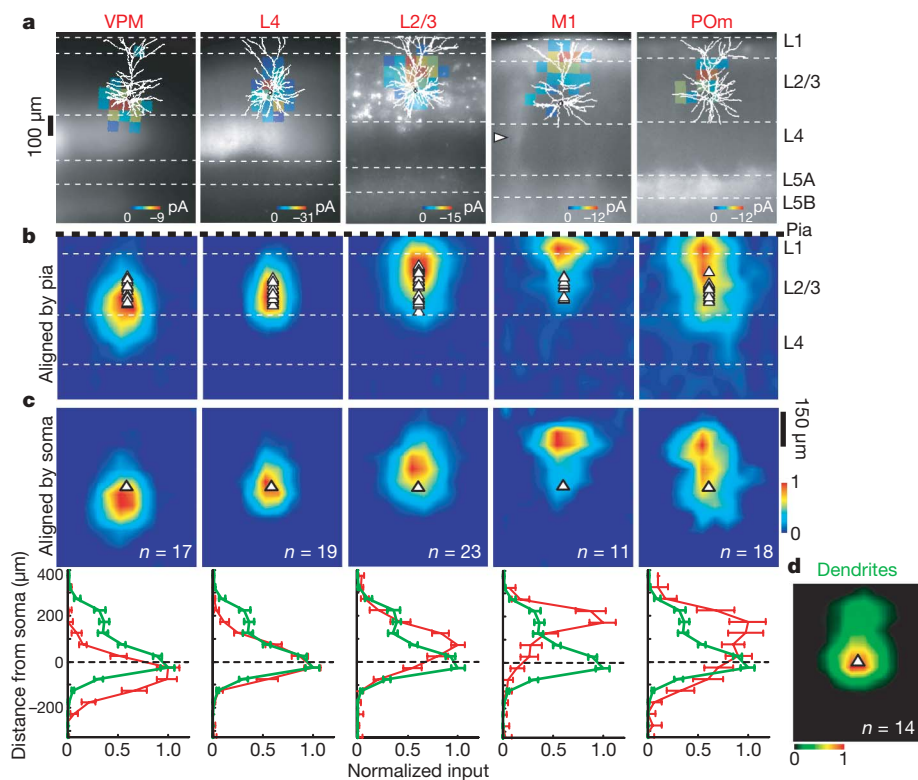

Figure 2 | Subcellular distribution of inputs onto L3 pyramidal neurons. a, Examples of sCRACM maps overlaid on reconstructed dendrites and fluorescence images showing ChR2-positive axons (VPM, M1 and POm) or axons and dendrites (L2/3 and L4). White arrowhead, bundle of ascending axons from M1. **b**, Group averages aligned by pia position (white triangles, soma position). **c**, Top: group averages aligned by soma position; bottom: vertical profiles of the distribution of synaptic input (red) and the dendritic length density (green; from **d**). Error bars, s.e.m. **d**, Average normalized dendritic length density of L3 pyramidal neurons.
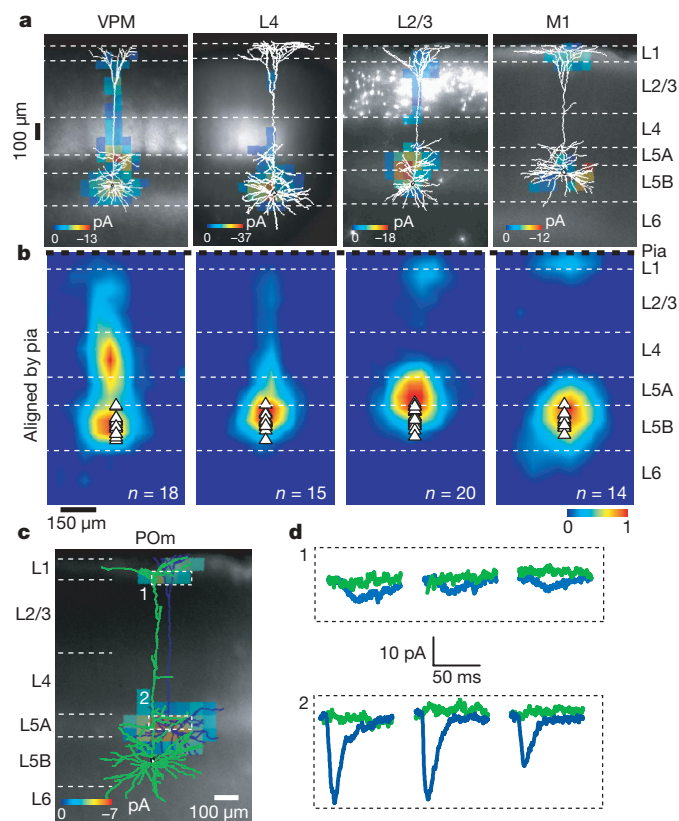
Figure 3 | Subcellular distribution of inputs onto L5B pyramidal neurons.
a, Examples of sCRACM maps overlaid on reconstructed dendritic
arborizations and fluorescence images. **b**, Group averages aligned by pia
position (white triangles, soma position). **c**, sCRACM map of POm input
onto a L5A pyramidal cell (blue). No responses were detected on the L5B
neuron (green). **d**, EPSC$_{sCRACM}$ recorded on the L5A neuron (blue) or the
L5B neuron (green) when photostimulating the boxed regions in **c** . The
stimulus occurred at the beginning of each trace.

of cells in the same column with identical laser powers (Supplementary
Table 1). L5B cells received 62-fold less input from POm in comparison
with L5A cells, despite there being pronounced overlap between L5B

dendrites and POm axons (dendritic length: L5A pyramidal cells
($n = 12$) in L1, 871 ± 546 μm, in L5A, 2,158 ± 899 μm; L5B pyramidal
cells ($n = 13$) in L1, 1,609 ± 732 μm, in L5A, 761 ± 350 μm; see Fig. 3c,
d, Supplementary Fig. 7 and Supplementary Table 1). L5A, but not
L5B, pyramidal cells received significant input from POm. This con-
firms that average cortical connectivity between populations of neu-
rons cannot always be deduced from the structure of axons and
dendrites alone[3–5]. Because L5B pyramidal neurons constitute the main
projection from barrel cortex to POm[19], there appears to be no dis-
ynaptic loop between these two areas.

By contrast with L3 cells, the inputs on L5 cells were not limited to a
single compartment, but were split into basal and apical domains
(Fig. 3, Supplementary Fig. 7), reinforcing the view that large pyr-
amidal neurons consist of multiple, weakly coupled compartments[20].
Here we describe the inputs to L5B neurons (Fig. 3), and then highlight
the differences with L5A neurons (Supplementary Fig. 7). VPM → L5B
input was distributed along most of the dendritic arborization, but was
most prominent on the basal dendrites and in L4. L4 → L5B input was
centred on the basal dendrites, overlapping with VPM input; weak
input was also detected along the apical dendrite up to the edge of
L1. L2/3 → L5B input was focused on the upper basal and apical
oblique dendrites, as well as on the apical tuft in L2. M1 → L5B input
was on the basal dendrites and on the apical tuft in L1.

Inputs to L5A pyramidal neurons similarly targeted dual dendritic
compartments (Supplementary Fig. 7), with some differences.
POm → L5A input was prominent, both on the basal dendrites and
the apical tuft in L1. Unlike for L2/3 → L5B input, L2/3 → L5A input
was centred on the basal dendrites (Fig. 4a, Supplementary Fig. 7).
For all L5 neurons taken together, there was a monotonic relation-
ship between the laminar position of the recorded cell and the loca-
tion of L2/3 input relative to the soma position (Fig. 4b–d). The
axodendritic overlap of L2/3 axons and L5 dendrites is most likely
an important factor in determining the subcellular location of L2/
3 → L5 input.

sCRACM maps functional neural circuits with subcellular resolu-
tion. Because sCRACM maps are based on somatic measurements of
synaptic currents generated in the dendrites, signal attenuation due
to dendritic filtering influences the structure of the maps. For
example, input on the apical tufts of L5 neurons could be reduced
several-fold relative to more proximal input[11]. sCRACM maps thus
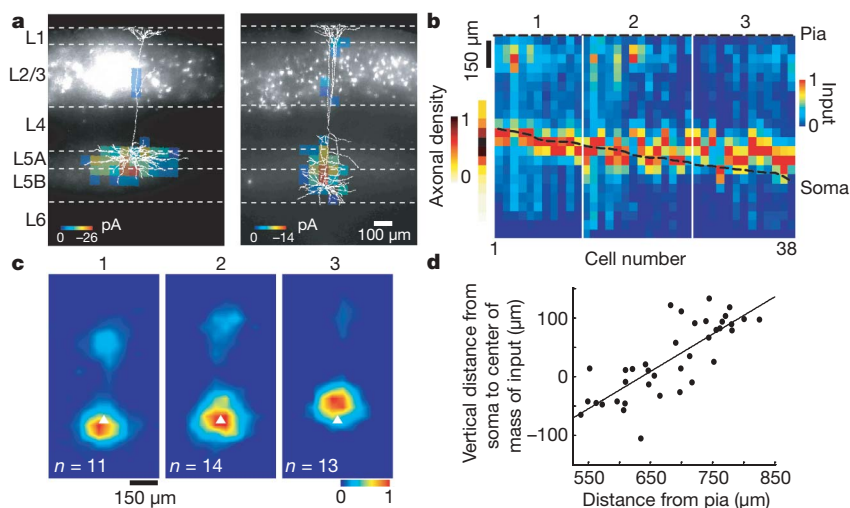represent a 'somatocentric' view of the dendritic distribution of

**Figure 4 | The laminar position of L5 pyramidal neurons determines the
dendritic location of L2/3 inputs. a**, Examples of the subcellular
distribution of L2/3 input onto superficial (L5A, left) or deep (L5B, right)
pyramidal neurons. **b**, Vertical profiles of the subcellular distribution of L2/
3 → L5 inputs. Each column represents one cell, ordered by cortical depth.
Cells were aligned by pia position. The relative density of L2/3 axons in the
deep layers is indicated to the left of the panel. **c**, Average subcellular location

of L2/3 input (aligned by soma position) of L5 cells grouped by increasing
distance from the pia (groups correspond to the white lines in **b**; white
triangles, soma position). **d**, Plot of the vertical distance from the soma to the
centre of mass of L2/3 input on the perisomatic region (<285 μm from the
soma) of L5 pyramidal neurons versus cortical depth. The line is a regression
fit.

synaptic input, where electrotonically distant synapses will appear relatively weak (Supplementary Fig. 8).

We mapped input from VPM, L4, L2/3, M1 and POm, within the dendritic arborizations of L3 and L5 neocortical pyramidal cells. L3 and L5 cells received input from most axonal populations, with the exception of POm→L5B. Some connections (VPM→L3, VPM→L5A, VPM→L5B, POm→L5A, L4→L3, L4→L5A, L2/3→L3, L2/3→L5A and L2/3→L5B) have been previously characterized[3,14,18,21,22], whereas others (M1→L3, M1→L5A, M1→L5B, POm→L3 and L4→L5B,) were previously unknown.

We identified several connections in L1: M1→L3, M1→L5A, M1→L5B, POm→L3 and POm→L5A. Axons from VPM, L2/3 and L4 did not contribute significantly to L1 input. POm neurons are thought to encode aspects of whisker position[23] and whisker M1 carries signals related to voluntary whisker control[24]. Our findings suggest that synapses in L1 carry signals related to whisker movement and position.

The spatial segregation of specific types of input within dendritic arborizations might subserve several functions. Segregated inputs are less likely to interact at the level of synaptic plasticity[25]. Spatially clustered co-active synapses are more efficacious in driving postsynaptic neurons than spatially distributed synapses[26,27]. For a fixed number of synapses, spatial segregation of different axonal populations within dendritic arborizations might thus strengthen the effective coupling between pre- and postsynaptic populations.

METHODS SUMMARY

Specific neuronal populations were labelled with ChR2 *in vivo* either by *in utero* electroporation[7,28] (L2/3 pyramidal cells), adeno-associated-virus infection (VPM, POm, M1) or infection with a Cre-recombinase-dependent adeno-associated virus[29] in mice expressing Cre (L4)[30]. Acute coronal slices were prepared from young adult (postnatal day 26–34) mice. Pyramidal neurons in L3 and L5 were recorded in a voltage clamp at room temperature (22–24 °C). Photostimulation was with blue (473-nm) laser pulses (duration, 1 ms; inter-stimulus interval, 400 ms; beam diameter, 6–16 μm) in the presence of TTX (1 μM), 3-((R)-2-carboxypiperazin-4-yl)-propyl-1-phosphonic acid (CPP, 5 μM) and 4-AP (100 μM).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Douglas, R. J. & Martin, K. A. Mapping the matrix: the ways of neocortex. *Neuron* **56**, 226–238 (2007).
2. London, M. & Hausser, M. Dendritic computation. *Annu. Rev. Neurosci.* **28**, 503–532 (2005).
3. White, E. L. Specificity of cortical synaptic connectivity: emphasis on perspectives gained from quantitative electron microscopy. *J. Neurocytol.* **31**, 195–202 (2002).
4. Dantzker, J. L. & Callaway, E. M. Laminar sources of synaptic input to cortical inhibitory interneurons and pyramidal neurons. *Nature Neurosci.* **3**, 701–707 (2000).
5. Shepherd, G. M. G., Stepanyants, A., Bureau, I., Chklovskii, D. B. & Svoboda, K. Geometric and functional organization of cortical circuits. *Nature Neurosci.* **8**, 782–790 (2005).
6. Nagel, G. *et al.* Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl Acad. Sci. USA* **100**, 13940–13945 (2003).
7. Petreanu, L., Huber, D., Sobczyk, A. & Svoboda, K. Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nature Neurosci.* **10**, 663–668 (2007).
8. Zhang, F., Wang, L. P., Boyden, E. S. & Deisseroth, K. Channelrhodopsin-2 and optical control of excitable cells. *Nature Methods* **3**, 785–792 (2006).
9. Larkum, M. E., Senn, W. & Luscher, H. R. Top-down dendritic input increases the gain of layer 5 pyramidal neurons. *Cereb. Cortex* **14**, 1059–1070 (2004).
10. Shu, Y., Yu, Y., Yang, J. & McCormick, D. A. Selective control of cortical axonal spikes by a slowly inactivating K+ current. *Proc. Natl Acad. Sci. USA* **104**, 11453–11458 (2007).
11. Williams, S. R. & Mitchell, S. J. Direct measurement of somatic voltage clamp errors in central neurons. *Nature Neurosci.* **11**, 790–798 (2008).
12. Binzegger, T., Douglas, R. J. & Martin, K. A. A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.* **24**, 8441–8453 (2004).
13. Gilbert, C. D. Microcircuitry of the visual cortex. *Annu. Rev. Neurosci.* **6**, 217–247 (1983).
14. Feldmeyer, D., Lubke, J. & Sakmann, B. Efficacy and connectivity of intracolumnar pairs of layer 2/3 pyramidal cells in the barrel cortex of juvenile rats. *J. Physiol. (Lond.)* **575**, 583–602 (2006).
15. Lu, S. M. & Lin, R. C. S. Thalamic afferents of the rat barrel cortex: a light- and electron-microscopic study using *Phaseolus vulgaris* leucoagglutinin as an anterograde tracer. *Somatosens. Mot. Res.* **10**, 1–16 (1993).
16. Koralek, K. A., Jensen, K. F. & Killackey, H. P. Evidence for two complementary patterns of thalamic input to the rat somatosensory cortex. *Brain Res.* **463**, 346–351 (1988).
17. Veinante, P. & Deschenes, M. Single-cell study of motor cortex projections to the barrel field in rats. *J. Comp. Neurol.* **464**, 98–103 (2003).
18. Lubke, J., Roth, A., Feldmeyer, D. & Sakmann, B. Morphometric analysis of the columnar innervation domain of neurons connecting layer 4 and layer 2/3 of juvenile rat barrel cortex. *Cereb. Cortex* **13**, 1051–1063 (2003).
19. Hoogland, P. V., Welker, E. & Van der Loos, H. Organization of the projections from barrel cortex to thalamus in mice studied with Phaseolus vulgaris-leucoagglutinin and HRP. *Exp. Brain Res.* **68**, 73–87 (1987).
20. Mainen, Z. F. & Sejnowski, T. J. Influence of dendritic structure on firing pattern in model neocortical neurons. *Nature* **382**, 363–366 (1996).
21. Bureau, I., von Saint Paul, F. & Svoboda, K. Interdigitated paralemniscal and lemniscal pathways in the mouse barrel cortex. *PLoS Biol.* **4**, e382 (2006).
22. Thomson, A. M. & Bannister, A. P. Interlaminar connections in the neocortex. *Cereb. Cortex* **13**, 5–14 (2003).
23. Yu, C., Derdikman, D., Haidarliu, S. & Ahissar, E. Parallel thalamic pathways for whisking and touch signals in the rat. *PLoS Biol.* **4**, e124 (2006).
24. Berg, R. W. & Kleinfeld, D. Vibrissa movement elicited by rhythmic electrical microstimulation to motor cortex in the aroused rat mimics exploratory whisking. *J. Neurophysiol.* **90**, 2950–2963 (2003).
25. Harvey, C. D. & Svoboda, K. Locally dynamic synaptic learning rules in pyramidal neuron dendrites. *Nature* **450**, 1195–1200 (2007).
26. Polsky, A., Mel, B. W. & Schiller, J. Computational subunits in thin dendrites of pyramidal cells. *Nature Neurosci.* **7**, 621–627 (2004).
27. Losonczy, A. & Magee, J. C. Integrative properties of radial oblique dendrites in hippocampal CA1 pyramidal neurons. *Neuron* **50**, 291–307 (2006).
28. Saito, T. & Nakatsuji, N. Efficient gene transfer into the embryonic mouse brain using *in vivo* electroporation. *Dev. Biol.* **240**, 237–246 (2001).
29. Atasoy, D., Aponte, Y., Su, H. H. & Sternson, S. M. A FLEX switch targets Channelrhodopsin-2 to multiple cell types for imaging and long-range circuit mapping. *J. Neurosci.* **28**, 7025–7030 (2008).
30. Liao, G. Y. & Xu, B. Cre recombinase-mediated gene deletion in layer 4 of murine sensory cortical areas. *Genesis* **46**, 289–293 (2008).

nature

METHODS

Electrophysiology and photostimulation. Neurons were patched with borosilicate pipettes (resistance, 4–6 MΩ). The intracellular solution contained 128 mM K-gluconate, 4 mM $MgCl_2$, 10 mM HEPES, 1 mM EGTA, 4 mM Na_2ATP, 0.4 mM Na_2GTP, 10 mM sodium phosphocreatine, 3 mM sodium L-ascorbate and 0.015 mM Alexa-594 (Molecular Probes) (pH 7.25, 290 mosM). Cells were recorded at depths of 50 to 95 μm in the brain slice. Data were acquired using custom programs (Ephus, available at https://openwiki.janelia.org/). Photostimulation was with a blue laser (473 nm; Crystal Laser). The beam's position was controlled with galvanometers (6210H, Cambridge Scanning). The beam was delivered through an air immersion objective (×4, 0.16 numerical aperture; UPlanApo, Olympus). The optics were designed to generate a nearly cylindrical beam (∼6–16 μm, full-width at half-maximum at the focal plane). The duration and intensity of the light pulses were controlled with a Pockels cell (ConOptics) and a shutter (LS3, Uniblitz).

For sCRACM mapping, we delivered light pulses (duration, 1 ms; inter-stimulus interval, 400 ms) on a 12 × 24 grid with 50-μm spacing (Fig. 1d). The grid area (0.6 × 1.2 mm²) included the entire thickness of the cortical grey matter. Stimuli were given in a spatial sequence pattern designed to maximize the time between stimuli to neighbouring spots[31]. To avoid sequence-specific responses during consecutive mapping, we flipped and rotated the stimulus pattern between maps. TTX (1 μM), CPP (5 μM) and 4-AP (100 μM) were added to the bath. Without 4-AP (or α-dendrotoxin (200 nM); Alomone Labs), TTX (1 μM) abolished 98 ± 1.9% of the excitatory postsynaptic currents evoked in the absence of drugs (six cells, 573 sites), even at high light intensities (>1 mW; Fig. 1b). When mapping inputs from L4 axons, we also added bicuculline (10 μM) to block contributions from GABAergic neurons in L4 (ref. 30). $EPSC_{sCRACM}$ were recorded in voltage clamp (−75 mV). Access resistances were <40 MΩ and stable (<20% change during the experiment); resting potentials were less than −55 mV.

$EPSC_{sCRACM}$ have relatively long delays (mean, 10.4 ± 2.5 ms; L2/3 → L5A perisomatic responses; <110 μm from the soma; 146 sites; 18 cells; Supplementary Fig. 2), most likely reflecting the slow charging and discharging of the axonal membrane (Supplementary Fig. 9). The delays varied across photostimulation sites (range, 6.4–21 ms) and $EPSC_{sCRACM}$ rise times (mean (10–90%), 6.5 ± 3.1 ms; range, 2.4–18.3 ms) and decay times (mean, 35 ± 28 ms; range, 6.2–160 ms) were long. $EPSC_{sCRACM}$ on occasion displayed multiple peaks. Minimal stimulation experiments (Supplementary Fig. 2c) revealed that unitary currents were slightly desynchronized at a single photostimulation location (latency jitter, 1.03 ± 0.5 ms; $n = 6$) and highly desynchronized across different locations (latency range, 10.4–17.2 ms; mean, 15.2 ± 3 ms; $n = 6$; Supplementary Fig. 2). The temporal smearing of the $EPSC_{sCRACM}$ waveform is therefore dominated by differences in latencies across different synapses. The rise time of the responses increased with distance to the soma, along both the apical and basal dendrites, consistent with filtering expected from cable theory (Supplementary Fig. 2)[32,33].

For each recorded cell, laser powers were adjusted to cause $EPSC_{sCRACM}$ with peak amplitudes of approximately −75 pA (L2/3 pyramidal neurons, −72 ± 47 pA; L5 pyramidal neurons, −84 ± 45 pA). The corresponding laser powers varied over one order of magnitude (120 μW–1.9 mW at the specimen plane), reflecting variations in the fraction of ChR2-positive axons and ChR2 expression levels across mice. sCRACM maps were repeated 2–5 times for each cell (Supplementary Fig. 4). After the recordings, dendritic arborizations were imaged using fluorescence microscopy (Retiga 2000RV, QImaging) and subsequently processed for biocytin staining and reconstructed. Only data from neurons where the apical dendrites ran parallel to the slice surface were included in the analysis.

Because photostimulation was with a cylindrical beam, sCRACM maps represent the two-dimensional projections of the three-dimensional distribution of inputs. As a consequence, the peak values of the distribution of inputs were sometimes centred on the somata, although somata are mostly devoid of excitatory synapses (Figs 2 and 3). This is analogous to the two-dimensional projection of the density of basal dendrites, which also peaks on the soma (Figs 2 and 3, Supplementary Fig. 7). Furthermore, under our conditions the sCRACM resolution was ∼60 μm, which is large in comparison with the diameters of most somata.

Data analysis. Individual pixels of sCRACM maps at position (x, y) ($Q(x, y)$) were computed as the mean $EPSC_{sCRACM}$ amplitude in a response window from 0 to 75 ms after the stimulus, and thus are a measure of charge. For consistency with previous studies, and because synaptic current is a more familiar unit, data are given in units of picoamperes. Q is given by

$$Q(x,y) \approx \rho_a(x,y)\rho_b(x,y)fqg(L(x,y))$$

where ρ_a and ρ_b are the densities of axons and dendrites, respectively; f is the filling fraction, defined as the fraction of axons making synapses with nearby dendrites[34]; q is the charge per synapse per light flash; and $g(L)$ is the dendritic attenuation as a function of electrotonic distance, $L(x, y)$, between the site of photostimulation and the soma. Because of dendritic attenuation, Q provides a somatocentric view of the synaptic input (see Supplementary Discussion). Because ρ_a, and possibly q, depends on details of the gene transfer method, it is challenging to compare the strengths of different projections onto the same cell. As ρ_a varies between preparations, sCRACM maps were normalized to the largest pixels within a map, and thus represent the relative strength of input within the dendritic tree.

Averaged $EPSC_{sCRACM}$ were scored as non-zero if their amplitudes (0–50 ms post-stimulus) were more than five times larger than the standard deviation of the baseline (Figs 1e, 2a, 3a and 4a, Supplementary Figs 3 and 7). Maps were either aligned on the soma or on the pia. In the case of alignment on the pia, it was necessary to correct for variations in cortical thickness; individual maps were therefore morphed by linear interpolation to a template based on the average cortical thickness. Similarly, the cortical depth of individual neurons (Fig. 4) was also normalized to the average depth across slices. Individual maps are presented as raw pixel images, whereas group averages are linearly interpolated without smoothing (for display only). To measure the density of L2/3 axons (Fig. 4b), we measured and peak-normalized mCherry fluorescence along the cortical axis in L4 to L6 in *in utero* electroporated animals ($n = 5$).

sCRACM resolution. The effective resolution of sCRACM mapping can be inferred from the point spread function. We measured the point spread function from the spatial distribution of the photostimulation sites that produce detectable responses in the vicinity of isolated dendritic branches. L5A cells often received input from L2/3 neurons along a single unbranched apical dendrite within L2/3 (Supplementary Fig. 3). To measure sCRACM resolution, we first identified the peak of L2/3 input on the apical dendrites of L5A cells within L2/3. We next photostimulated in a line across the apical dendrite, through the peak of L2/3 input (12 positions, 15-μm spacing between stimuli, inter-stimulus interval of 6 s; Supplementary Fig. 3a). Because the activated synapses were on a single dendritic branch in the vicinity of the photostimuli, the spatial distribution of responses represents a measure of the spatial resolution. After the experiment, the dendritic arborizations of the recorded neurons were reconstructed. Only cells where the apical dendrite did not ramify within 100 μm of the photostimulation sites were included for analysis. For the light intensities used for sCRACM mapping, the full-width at half-maximum of the spatial profile of the responses was 59 ± 14 μm ($n = 4$; Supplementary Fig. 3b, c). Higher laser intensities degraded the resolution slightly.

To verify this resolution estimate within our data set, we identified stretches of unbranched L5 apical dendrites that received input from ChR2-positive axons originating either in VPM, L4 or L2/3. Only inputs separated from branch points by at least 100 μm were used. In addition, to avoid over-representation of a subset of inputs, we only scored inputs that were at least 100 μm apart. Detectable $EPSC_{sCRACM}$ were only evoked within 75 μm of the apical dendrite (Supplementary Fig. 3d, e). We conclude that sCRACM maps specific types of input with ∼60-μm spatial resolution.

Data in the text are given as mean ± s.d.

31. Shepherd, G. M., Pologruto, T. A. & Svoboda, K. Circuit analysis of experience-dependent plasticity in the developing rat barrel cortex. *Neuron* **38**, 277–289 (2003).
32. Williams, S. R. & Stuart, G. J. Site independence of EPSP time course is mediated by dendritic I(h) in neocortical pyramidal neurons. *J. Neurophysiol.* **83**, 3177–3182 (2000).
33. Nevian, T., Larkum, M. E., Polsky, A. & Schiller, J. Properties of basal dendrites of layer 5 pyramidal neurons: a direct patch-clamp recording study. *Nature Neurosci.* **10**, 206–214 (2007).
34. Stepanyants, A., Hof, P. R. & Chklovskii, D. B. Geometry and structural plasticity of synaptic connectivity. *Neuron* **34**, 275–288 (2002).

LETTERS

Deficiency of a β-arrestin-2 signal complex contributes to insulin resistance

Bing Luan[1], Jian Zhao[1], Haiya Wu[3], Baoyu Duan[1], Guangwen Shu[1], Xiaoying Wang[4], Dangsheng Li[2], Weiping Jia[3], Jiuhong Kang[1] & Gang Pei[1,5]

Insulin resistance, a hallmark of type 2 diabetes, is a defect of insulin in stimulating insulin receptor signalling[1,2], which has become one of the most serious public health threats. Upon stimulation by insulin, insulin receptor recruits and phosphorylates insulin receptor substrate proteins[3], leading to activation of the phosphatidylinositol-3-OH kinase (PI(3)K)–Akt pathway. Activated Akt phosphorylates downstream kinases and transcription factors, thus mediating most of the metabolic actions of insulin[4–6]. β-arrestins mediate biological functions of G-protein-coupled receptors by linking activated receptors with distinct sets of accessory and effecter proteins, thereby determining the specificity, efficiency and capacity of signals[7–11]. Here we show that in diabetic mouse models, β-arrestin-2 is severely downregulated. Knockdown of β-arrestin-2 exacerbates insulin resistance, whereas administration of β-arrestin-2 restores insulin sensitivity in mice. Further investigation reveals that insulin stimulates the formation of a new β-arrestin-2 signal complex, in which β-arrestin-2 scaffolds Akt and Src to insulin receptor. Loss or dysfunction of β-arrestin-2 results in deficiency of this signal complex and disturbance of insulin signalling _in vivo_, thereby contributing to the development of insulin resistance and progression of type 2 diabetes. Our findings provide new insight into the molecular pathogenesis of insulin resistance, and implicate new preventive and therapeutic strategies against insulin resistance and type 2 diabetes.

We first investigated the expression pattern of β-arrestins in the _db/db_ mouse model of type 2 diabetes. There was a decrease in β-arrestin-2 protein and messenger RNA levels in liver and skeletal muscle of _db/db_ mice (Fig. 1a, b). Consistently, in dietary-induced insulin-resistant model mice (high-fat diet (HFD) mice)[12], expression of β-arrestin-2 also decreased in liver and skeletal muscle (Fig. 1c, d). Expression levels of β-arrestin-2 in adipose tissue (Fig. 1), brain or lung of _db/db_ or HFD mice did not change (data not shown). β-arrestin-1 was downregulated in liver and skeletal muscle, but to a lesser extent than β-arrestin-2 (data not shown). We also found a similar downregulation of β-arrestins in liver from clinical samples of type 2 diabetes (eight pairs) (Supplementary Fig. 1). Changes in β-arrestin levels in these insulin-resistant mice and clinical samples raise the possibility that β-arrestins might have a role in insulin signalling and that its deficiency might contribute to insulin resistance and type 2 diabetes.

We then assessed the potential roles of β-arrestin-2 in whole-body insulin action and glucose metabolism. β-arrestin-2 knockout (β-arr2-KO) mice are viable and fertile, with normal body weight and food intake (Supplementary Fig. 2a). Their fasted blood glucose and insulin levels were normal, but re-fed blood glucose and insulin

levels were much higher than wild-type littermates (Fig. 2a, b), suggesting a potential decrease in systemic insulin sensitivity. In fact, glucose tolerance tests (GTTs) revealed significant deterioration in glucose metabolism in β-arr2-KO mice (Fig. 2c). Insulin secretion in response to glucose load during GTTs was also higher in β-arr2-KO mice (Supplementary Fig. 2c), consistent with insulin tolerance tests (ITTs) which showed significantly decreased insulin sensitivity in β-arr2-KO mice (Fig. 2d). To define the role for β-arrestin-2 in modulating whole-body insulin sensitivity further, we performed hyperinsulinaemic–euglycaemic clamp studies. We found no significant difference in basal hepatic glucose production between β-arr2-KO mice and wild-type littermates, but clamp hepatic glucose production was significantly higher in β-arr2-KO mice (Supplementary Fig. 2d). Whole-body glucose disposal and infusion rates were significantly decreased in β-arr2-KO mice (Supplementary Fig. 2e). These results together demonstrate that β-arr2-KO mice exhibited impaired insulin sensitivity.
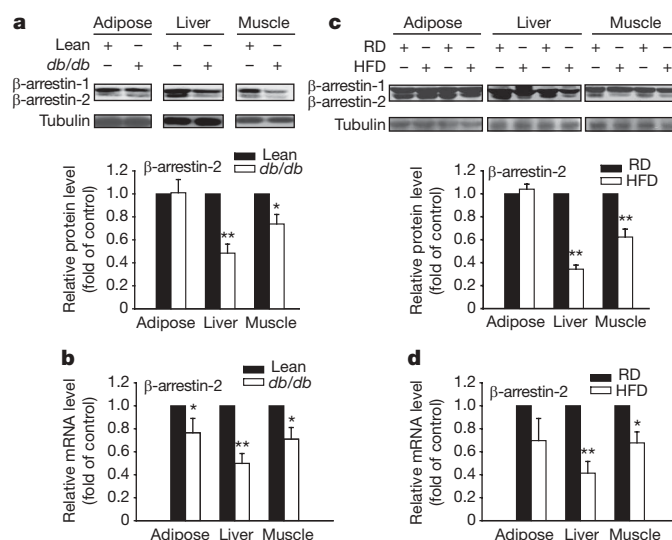

Figure 1 | Downregulation of β-arrestin-2 in diabetic mice. a, b, Immunoblot (**a**) and quantitative RT–PCR (**b**) of β-arrestin expression in adipose tissue, liver and skeletal muscle of lean (_n_ = 5) and _db/db_ mice (_n_ = 5). Densitometric analysis is shown. **c, d,** Immunoblot (**c**) and quantitative PCR with reverse transcription (RT–PCR) (**d**) of β-arrestin expression in adipose tissue, liver and skeletal muscle of C57BL/6 mice fed with a regular diet (RD) (_n_ = 5) or HFD (_n_ = 5). Densitometric analysis is shown. Data are presented as mean and s.e.m. *_P_ < 0.05, **_P_ < 0.005, versus control.

[1]Laboratory of Molecular Cell Biology, Institute of Biochemistry and Cell Biology, and Graduate School of the Chinese Academy of Sciences, [2]Shanghai Information Center for Life Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 200031, Shanghai, China. [3]Department of Endocrinology and Metabolism, Shanghai Jiaotong University Affiliated Sixth People's Hospital; Shanghai Diabetes Institute; Shanghai Clinical Center of Diabetes, 200233, Shanghai, China. [4]Fudan University Affiliated Zhongshan Hospital, 200032, Shanghai, China. [5]School of Life Science and Technology, Tongji University, 200092, Shanghai, China.
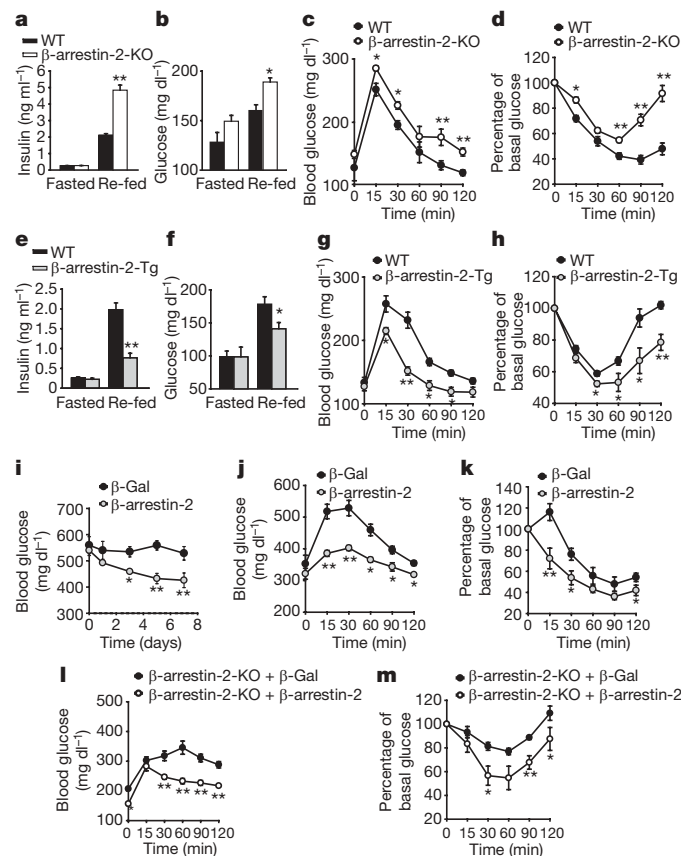
Figure 2 | β-arrestin-2 affects the development of insulin resistance.
a, b, Serum insulin (**a**) and blood glucose (**b**) levels in β-arr2-KO mice
($n = 8$) and wild-type (WT) littermates ($n = 8$) under fasted and re-fed
stages. **c, d,** Glucose levels during GTTs (1 g kg^{-1}) (**c**) and ITTs (1 U kg^{-1})
(**d**) in β-arr2-KO mice ($n = 10$) and wild-type littermates ($n = 8$). **e, f,** Serum
insulin (**e**) and blood glucose (**f**) levels in β-arr2-Tg mice ($n = 8$) and wild-
type littermates ($n = 8$) under fasted and re-fed stages. **g, h,** GTTs (1 g kg^{-1})
(**g**) and ITTs (0.75 U kg^{-1}) (**h**) in β-arr2-Tg mice ($n = 11$) and wild-type
littermates ($n = 8$). **i,** Blood glucose concentration of *db/db* mice ($n = 8$)
injected with indicated adenovirus. **j, k,** GTTs (1.5 g kg^{-1}) (**j**) and ITTs
(1.5 U kg^{-1}) (**k**) in *db/db* mice at 7 days after adenovirus injection. **l, m,** GTTs
(1.5 g kg^{-1}) (**l**) and ITTs (1.5 U kg^{-1}) (**m**) in β-arr2-KO mice fed on HFD at
7 days after adenovirus injection. Data are presented as mean ± s.e.m.
*$P < 0.05$, **$P < 0.005$, versus control.

Transgenic mice expressing human β-arrestin-2 driven by the
cytomegalovirus (CMV) promoter (β-arr2-Tg) had an approximate
twofold increase in liver β-arrestin-2 expression compared with con-
trol mice (Supplementary Fig. 2b). β-arr2-Tg mice develop normally
with normal food intake and body weight (Supplementary Fig. 2b).
Their serum insulin and blood glucose levels were similar to wild-
type mice in fasted stages but much lower in re-fed stages (Fig. 2e, f).
GTTs and ITTs showed augmented glucose metabolism and insulin
sensitivity in β-arr2-Tg mice (Fig. 2g, h, and Supplementary Fig. 2c).
Similar results were obtained using β-arr2-KO, β-arr2-Tg and their
wild-type littermates that had been fed on HFD (Supplementary Fig.
3), further suggesting a potential anti-diabetic role of β-arrestin-2.

We next explored the possibility that administration of β-arrestin-2
may have a therapeutic potential against insulin resistance and type 2
diabetes. To this end, we administered β-arrestin-2 in *db/db* mice by
using adenovirus. Intravenous injection of recombinant adenovirus
expressing β-arrestin-2 led to an approximate threefold increase of
β-arrestin-2 protein levels in liver, without altering the food intake
and body weight (Supplementary Fig. 2f, g). The blood glucose levels
under fed conditions were reduced in *db/db* mice that received the
β-arrestin-2 adenovirus, compared with control mice (Fig. 2i).
Injection of β-arrestin-2 adenovirus ameliorated glucose tolerance

and insulin sensitivity, as shown in GTTs and ITTs (Fig. 2j, k).
Similarly, rescue of β-arrestin-2 expression by adenovirus injection
in β-arr2-KO mice fed on HFD amended glucose intolerance and
insulin resistance in these mice (Fig. 2l, m). These results indicate
that β-arrestin-2 is a positive regulator of insulin sensitivity.

At the molecular level, insulin resistance results from defects in
insulin signalling in peripheral tissues[4]. Interestingly, although activa-
tion of PI(3)K by insulin was similar in mice of all genotypes
(Supplementary Fig. 4a), we observed different activities of Akt in liver
of β-arr2-KO, β-arr2-Tg and wild-type mice, as monitored by *in vitro*
kinase assay (Supplementary Fig. 4b). Further, insulin-stimulated
phosphorylation of Akt, GSK3-β and Foxo1 were dramatically
reduced in liver of β-arr2-KO mice but largely increased in β-arr2-Tg
mice (Fig. 3a, b, and Supplementary Fig. 4c). Similar results were also
observed in skeletal muscle and adipose tissue (Supplementary Fig.
4d–g). These results show that β-arrestin-2 promotes insulin-
stimulated activation of Akt but does not affect PI(3)K activity.

Recent studies have shown that phosphorylation of Akt at Tyr 315/
326 by Src enhances Akt serine/threonine phosphorylation and is a
prerequisite of full Akt activation[13–17]. We have also observed that, in
the presence of Src inhibitor PP2, tyrosine phosphorylation and serine/
threonine phosphorylation of Akt were remarkably reduced in Hep3B
hepatocytes (Supplementary Fig. 5a). Furthermore, Akt activity was
reduced when Tyr 315/326 were substituted by Ala (Supplementary
Fig. 5b). In mouse liver, we observed that insulin stimulation led to an
approximate twofold increase in tyrosine phosphorylation of Akt in
wild-type mice (Fig. 3c and Supplementary Fig. 5c), which was
reduced by about 40% in β-arr2-KO mice and increased by about
1.5-fold in β-arr2-Tg mice (Fig. 3c and Supplementary Fig. 5c).
Interestingly, knockdown of Src by short interfering RNA (siRNA)
(Fig. 3d) in primary hepatocytes isolated from β-arr2-KO mice and
their wild-type littermates abolished the difference in Akt activities
(Fig. 3e) and moderately depressed the differential reduction on
g6p/pepck expression (Supplementary Fig. 5d, e). The suppressive
effect on hepatocytic glucose production conferred by β-arrestin-2
was also abolished in the presence of Src siRNA (Fig. 3f). Together,
these results suggest that β-arrestin-2 promotes Akt activation and
glucose metabolism through Src.

β-arrestins have been reported to function as adaptors and to
promote the activation of various mitogen-activated protein kinases
such as ERK1/2 and JNK3 (ref. 18). A similar scaffolding mechanism
might also operate in insulin signalling. We found that in liver of
C57BL/6 mice, endogenous β-arrestin-2, Akt and Src were co-purified
with each other (Fig. 3g). However, the association between Src and
Akt was dramatically reduced in liver samples from β-arr2-KO mice
(Fig. 3h and Supplementary Fig. 6a). Conversely, Akt and Src inter-
action was remarkably enhanced in β-arr2-Tg mice (Fig. 3i and
Supplementary Fig. 6b). These results strongly suggest that
β-arrestin-2, Akt and Src form a complex *in vivo* and that
β-arrestin-2 is essential for mediating the association of Akt with
Src. Next, we examined whether the formation of this Akt/β-
arrestin-2/Src signalling complex depends on insulin stimulation.
Intraperitoneal administration of insulin triggered a marked increase
in Akt/Src interaction as well as Akt/β-arrestin-2 association in livers of
C57BL/6 mice (Fig. 3j and Supplementary Fig. 6c), implicating the
promotion of Akt/β-arrestin-2/Src interaction by insulin stimulation.
Furthermore, we found β-arrestin-2 interacts directly with insulin
receptor (Supplementary Fig. 6d). β-arrestin-2 as well as Akt and Src
associate with insulin receptor in a similar time-dependent manner
(Fig. 3k and Supplementary Fig. 6e). Moreover, interaction of Akt and
Src with insulin receptor depends on the expression level of β-arrestin-2
(Supplementary Fig. 6f, g), supporting the idea that β-arrestin-2 is
essential for the association of Akt/Src with insulin receptor.

We examined a series of truncation mutants of β-arrestin-2 by
immunoprecipitation assay. A β-arrestin-2 fragment comprising
amino acids 1–185 interacts with Src as efficiently as wild-type
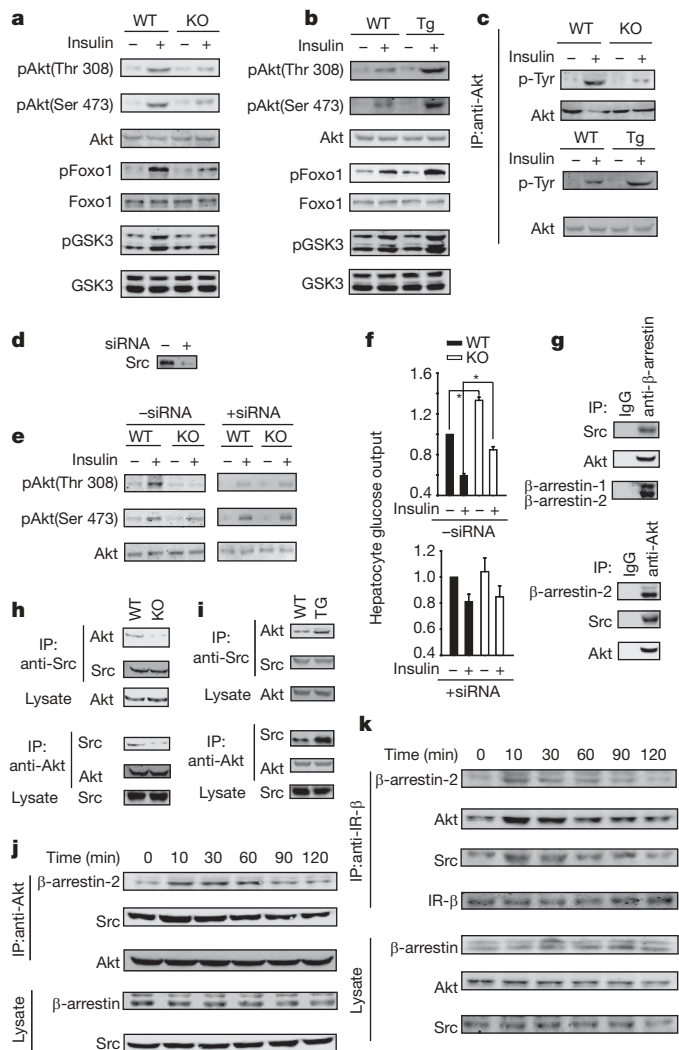β-arrestin-2 but completely loses the interaction with Akt (Fig. 4a).

Figure 3 | Insulin stimulated the formation of Receptor/Akt/β-arrestin-2/Src signal complex. a, b, Activation of Akt in livers of β-arr2-KO mice (n = 6) (**a**), β-arr2-Tg mice (n = 6) (**b**) and their wild-type littermates (n = 6). Mice were injected with either saline or insulin (1 U kg⁻¹) for 10 min. **c,** Tyrosine phosphorylation of Akt immunopurified from livers of β-arr2-KO mice (n = 6), β-arr2-Tg mice (n = 6), and their wild-type littermates (n = 6). **d,** Suppression of Src in primary hepatocytes by Src siRNA. **e,** Akt activities in primary hepatocytes shown in **d** (n = 3). **f,** Glucose production treated with or without 10 nM insulin for 3 h in primary hepatocytes shown in **d** (n = 3). **g,** Interactions of Akt/β-arrestin-2/Src assayed by immunoprecipitation from livers of C57BL/6 mice (n = 5). **h, i,** Interaction of Akt and Src in liver extracts of β-arr2-KO mice (n = 6) (**h**), β-arr2-Tg mice (n = 6) (**i**), and their wild-type littermates (n = 6) were assayed by immunoprecipitation. **j,** Insulin induces formation of Akt/β-arrestin-2/Src complex. Immunoprecipitation was conducted in liver extracts of C57BL/6 mice (n = 6) injected with 1 U kg⁻¹ insulin for the indicated times. **k,** Insulin stimulated interaction of Akt/β-arrestin-2/Src with insulin receptor (n = 6). Data are presented as mean and s.e.m. *P < 0.05, **P < 0.005, versus control.

Overexpression of β-arrestin-2 1–185 in Hep3B hepatocytes inhibited Akt/Src interaction (data not shown) and dramatically reduced insulin-stimulated Akt activation (Fig. 4b), presumably by competing with endogenous β-arrestin-2. Moreover, overexpression of β-arrestin-2 186–409, a truncated mutant that interacts with insulin receptor but not with Akt or Src (Fig. 4c) in Hep3B hepatocytes, dominant-negatively suppressed insulin-stimulated Akt activation (Fig. 4d). These results clearly demonstrate the pivotal role of β-arrestin-2 in scaffolding the active insulin receptor/Akt/β-arrestin-2/Src signalling complex after insulin stimulation.
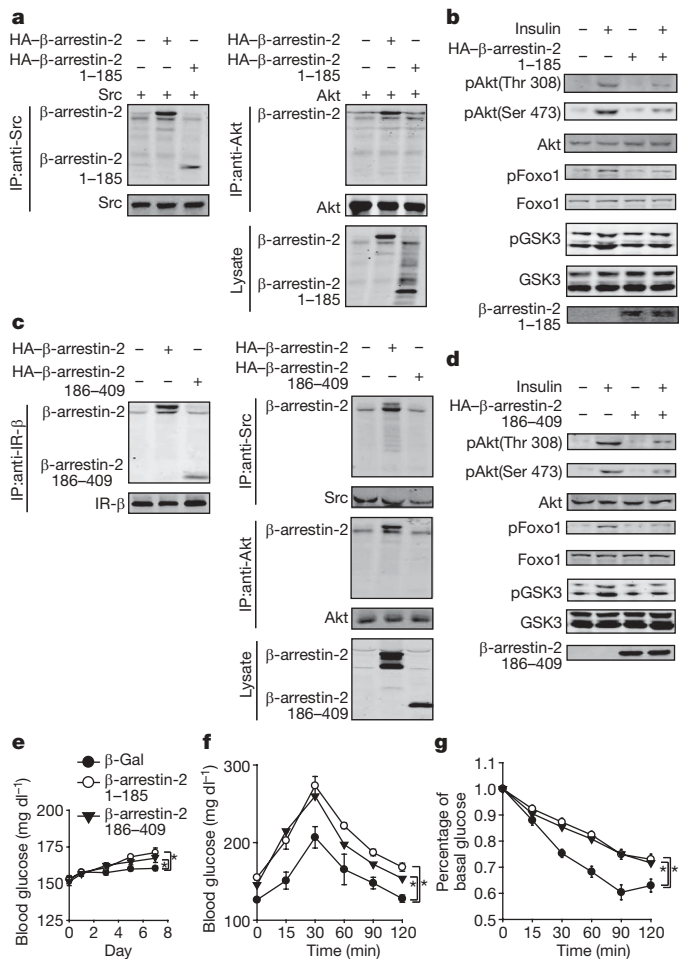
Figure 4 | Mutation of β-arrestin-2 contributes to insulin resistance in vivo. a, β-arrestin-2 1–185 interacts with Src but not Akt. Immunoprecipitation was conducted from Hep3B hepatocytes transfected with indicated plasmids (n = 3). **b,** β-arrestin-2 1–185 inhibits activation of Akt in Hep3B hepatocytes (n = 3). **c,** β-arrestin-2 186–409 interacts with insulin receptor-β but not Akt or Src in Hep3B hepatocytes (n = 3). **d,** β-arrestin-2 186–409 inhibits activation of Akt in Hep3B hepatocytes (n = 3). **e,** Blood glucose concentration of C57BL/6 mice (n = 8) injected with indicated adenovirus. **f, g,** GTTs (1 g kg⁻¹) (**f**) and ITTs (1 U kg⁻¹) (**g**) in C57BL/6 mice at 7 days after adenovirus injection. Data are presented as mean ± s.e.m. *P < 0.05, **P < 0.005, versus control.

We further tested whether expression of these β-arrestin-2 mutants would contribute to insulin resistance in vivo. Adenoviruses encoding β-Gal, β-arrestin-2 1–185 or β-arrestin-2 186–409 were injected intravenously into C57BL/6 or db/db diabetic mice. C57BL/6 mice infected with adenoviruses expressing β-arrestin-2 1–185 and β-arrestin-2 186–409 showed higher blood glucose levels than those of control mice (Fig. 4e). β-arrestin-2 1–185 and 186–409 adenovirus injection also deteriorated glucose tolerance and insulin sensitivity, as shown in GTTs and ITTs (Fig. 4f, g). Similar results were obtained in db/db diabetic mice (Supplementary Fig. 7).

Taken together, we have shown that assignment of Akt with Src and activated insulin receptor by β-arrestin-2 is essential for proper insulin signalling and whole-body insulin action. In contrast to the classical known pathway, β-arrestin-2 mediates activation of Akt through Src, which does not affect PI(3)K. However, this β-arrestin-2-Src pathway and PI(3)K pathway are not necessarily separated. One possibility could be that insulin receptor substrates and PI(3)K, as well as PDKs, may be involved in the signal complex, integrating these two pathways into one insulin-signalling network for efficient signal transduction (Supplementary Fig. 7d). Current and future investigations of the function and mechanism of this

novel signal complex will provide new insight into the understanding of insulin resistance and type 2 diabetes, and uncover potential molecular targets for treating metabolic diseases.

METHODS SUMMARY

Immunoprecipitation and immunoblotting. Mouse tissues were quickly excised and frozen in liquid nitrogen. Tissue lysate was prepared and used for immunoprecipitation and immunoblotting as described[19,20]. Blots were incubated with IRDyeTM800CW-conjugated secondary antibody. The image was captured and analysed by the Odyssey infrared imaging system and Scion Image (Li-Cor Bioscience).

mRNA analysis. We analysed β-arrestin mRNA levels by real-time PCR after reverse transcription as described[21]. Hypoxanthine–guanine phosphoribosyl-transferase (HPRT) mRNA levels were used for normalization between samples.

GTTs and ITTs. For GTTs, mice were injected intraperitoneally with glucose after starvation for 6 h. Blood glucose was measured at indicated time points. We injected insulin intraperitoneally to mice under fed conditions for ITTs. We collected blood and determined the glycaemia using a glucometer (Roche Accu-chek). Serum insulin levels were measured using a rat/mouse enzyme-linked immunosorbent assay (ELISA) kit.

Statistical analysis. All data are shown as mean ± s.e.m. Measurements at single time points were analysed by analysis of variance (ANOVA) or, if appropriate, by Student's t-test. Time courses were analysed by repeated-measurements (mixed-model) ANOVA with Bonferroni post-tests.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Matthaei, S., Stumvoll, M., Kellerer, M. & Haring, H. U. Pathophysiology and pharmacological treatment of insulin resistance. *Endocr. Rev.* **21,** 585–618 (2000).
2. Taniguchi, C. M., Emanuelli, B. & Kahn, C. R. Critical nodes in signalling pathways: insights into insulin action. *Nature Rev. Mol. Cell Biol.* **7,** 85–96 (2006).
3. Sun, X. J. *et al.* Structure of the insulin receptor substrate IRS-1 defines a unique signal transduction protein. *Nature* **352,** 73–77 (1991).
4. Biddinger, S. B. & Kahn, C. R. From mice to men: insights into the insulin resistance syndromes. *Annu. Rev. Physiol.* **68,** 123–158 (2006).
5. Franke, T. F. *et al.* The protein kinase encoded by the Akt proto-oncogene is a target of the PDGF-activated phosphatidylinositol 3-kinase. *Cell* **81,** 727–736 (1995).
6. Burgering, B. M. & Coffer, P. J. Protein kinase B (c-Akt) in phosphatidylinositol-3-OH kinase signal transduction. *Nature* **376,** 599–602 (1995).
7. McDonald, P. H. *et al.* β-arrestin 2: a receptor-regulated MAPK scaffold for the activation of JNK3. *Science* **290,** 1574–1577 (2000).
8. Luttrell, L. M. *et al.* Activation and targeting of extracellular signal-regulated kinases by beta-arrestin scaffolds. *Proc. Natl Acad. Sci. USA* **98,** 2449–2454 (2001).
9. Luttrell, L. M. *et al.* β-arrestin-dependent formation of β2 adrenergic receptor-Src protein kinase complexes. *Science* **283,** 655–661 (1999).
10. Beaulieu, J. M. *et al.* An Akt/β-arrestin 2/PP2A signaling complex mediates dopaminergic neurotransmission and behavior. *Cell* **122,** 261–273 (2005).
11. Beaulieu, J. M. *et al.* A β-arrestin 2 signaling complex mediates lithium action on behavior. *Cell* **132,** 125–136 (2008).
12. Yang, Q. *et al.* Serum retinol binding protein 4 contributes to insulin resistance in obesity and type 2 diabetes. *Nature* **436,** 356–362 (2005).
13. Chen, R. *et al.* Regulation of Akt/PKB activation by tyrosine phosphorylation. *J. Biol. Chem.* **276,** 31858–31862 (2001).
14. Jiang, T. & Qiu, Y. Interaction between Src and a C-terminal proline-rich motif of Akt is required for Akt activation. *J. Biol. Chem.* **278,** 15789–15793 (2003).
15. Craxton, A., Jiang, A., Kurosaki, T. & Clark, E. A. Syk and Bruton's tyrosine kinase are required for B cell antigen receptor-mediated activation of the kinase Akt. *J. Biol. Chem.* **274,** 30644–30650 (1999).
16. Wong, B. R. *et al.* TRANCE, a TNF family member, activates Akt/PKB through a signaling complex involving TRAF6 and c-Src. *Mol. Cell* **4,** 1041–1049 (1999).
17. Datta, K., Bellacosa, A., Chan, T. O. & Tsichlis, P. N. Akt is a direct target of the phosphatidylinositol 3-kinase. Activation by growth factors, v-src and v-Ha-ras, in Sf9 and mammalian cells. *J. Biol. Chem.* **271,** 30835–30839 (1996).
18. DeWire, S. M., Ahn, S., Lefkowitz, R. J. & Shenoy, S. K. β-arrestins and cell signaling. *Annu. Rev. Physiol.* **69,** 483–510 (2007).
19. Gao, H. *et al.* Identification of β-arrestin2 as a G protein-coupled receptor-stimulated regulator of NF-κB pathways. *Mol. Cell* **14,** 303–317 (2004).
20. Luan, B., Zhang, Z., Wu, Y., Kang, J. & Pei, G. β-arrestin2 functions as a phosphorylation-regulated suppressor of UV-induced NF-κB activation. *EMBO J.* **24,** 4237–4246 (2005).
21. Kang, J. *et al.* A nuclear function of β-arrestin1 in GPCR signaling: regulation of histone acetylation and gene transcription. *Cell* **123,** 833–847 (2005).

METHODS

Mice. β-arr2-KO mice were provided by R. J. Lefkowitz (Duke University Medical Center). β-arr2-Tg mice were generated as described[22]. All other mice were from Shanghai Laboratory Animal Center, Chinese Academy of Sciences. Mice were fed with a regular diet (Formulab 5008, Labdiet 5053) or high-fat diet (55% fat calories) (Harlan-Teklad 93075) and had free access to water and diet. We injected adenovirus (7×10^9 viral particles per 100 μl saline) into the tail vein of db/db and C57BL/6 mive specifically to target the liver. Body weight and food intake were measured as described[23]. Lean mice were 8 weeks old and fed with a regular diet (body weight 20.3 ± 1.2 g, food intake 3.5 ± 0.5 g), db/db mice were 8 weeks old and fed with a regular diet (body weight 38.5 ± 1.4 g, food intake 4.8 ± 0.7 g). Six-week-old C57BL/6 mice were fed on regular diet (body weight 22.6 ± 2.1 g, food intake 3.6 ± 0.3 g) or HFD (body weight 36.7 ± 1.9 g, food intake 4.1 ± 0.2 g) for an additional 8 weeks. All mice except for HFD feeding were 8 weeks old when experiments were conducted; db/db mice were 8 weeks old when they received adenovirus injection. GTTs and ITTs were conducted at 7 days after adenovirus injection.

Cell transfection and plasmids. Human hepatocyte Hep3B cells were transfected by LipofectAMINE (Invitrogen). For all transfection experiments, CMV-β-Gal was used to compensate the total DNA input. Full lengths of Akt and Src were cloned into modified pcDNA3 vector in-frame with HA or Flag at the amino (N) terminus. Akt Tyr315/326Ala was also cloned into modified pcDNA3 vector in-frame with Flag at the N terminus. Plasmids containing complementary DNA (cDNA) encoding β-arrestin-2 and its truncation mutants were generated as described[19]. The authenticity of the DNA sequences was confirmed by sequencing.

Materials and reagents. Anti-β-arrestin rabbit polyclonal antibody (A1CT and A2CT) was a gift from R. J. Lefkowitz (Duke University Medical Center). Antibodies directed against Akt (total), Akt (phosphorylated Thr308), Akt (phosphorylated Ser473), GSK3-α/β (phosphorylated Ser21/9), Foxo1 (total), Foxo1 (phosphorylated Ser256), Src, phosphorylated Tyr and insulin receptor were obtained from Cell Signalling. GSK3-α/β antibody was from Santa Cruz. PP2 and wortmannin were from Sigma. The Rat/Mouse Insulin ELISA Kit was from Linco Research. The PI3-Kinase ELISA Kit was from Echelon Biosciences. The Rat/Mouse Glucagon ELISA Kit, NEFA, triglycerides, and cholesterol detection kit were from WAKO Chemicals USA. The Mouse Epinephrine ELISA Kit was from USCN Life Science. The Akt Kinase Assay Kit was from Cell Signalling.

Adenovirus preparation and injection. We generated adenoviruses encoding Gal, β-arrestin-2, β-arrestin-2 1–185 and β-arrestin-2 186–409 using the adEasy system according to the manufacturer's instructions (Stratagene).

Quantitative PCR coupled with RT–PCR. The primer pairs used were: mouse β-arrestin-1 sense, 5′- AAGGGACACGAGTGTTCAAGA-3′; antisense, 5′-CCC GCT TTC CCA GGT AGA C-3′; mouse β-arrestin-2 sense, 5′-GGC AAG CGC GAC TTT GTA G-3′; antisense, 5′- GTG AGG GTC ACG AAC ACT TTC-3′; mouse HPRT sense, 5′-CCT GCT GGA TTA CAT TAA AGC ACT G-3′; antisense, 5′-TTC AAC ACT TCG AGA GGT CCT-3′.

Primary hepatocyte culture and glucose production. Primary hepatocytes were isolated and cultured after perfusion and collagenase digestion of the liver[24]. Glucose production in primary hepatocytes was measured as described[12], and a Glucose (GO) Assay Kit from Sigma was used to detect glucose concentration.

Hyperinsulinaemic–euglycaemic clamps. Hyperinsulinaemic–euglycaemic clamps were performed in 8-week-old mice as described[25,26]. Insulin was infused at $2.5 \, \text{mU kg}^{-1} \, \text{min}^{-1}$.

22. Shi, Y. et al. Critical regulation of CD4[+] T cell survival and autoimmunity by β-arrestin 1. Nature Immunol. **8**, 817–824 (2007).

23. Netea, M. G. et al. Deficiency of interleukin-18 in mice leads to hyperphagia, obesity and insulin resistance. Nature Med. **12**, 650–656 (2006).

24. Arkan, M. C. et al. IKK-β links inflammation to obesity-induced insulin resistance. Nature Med. **11**, 191–198 (2005).

25. Shen, H. Q., Zhu, J. S. & Baron, A. D. Dose-response relationship of insulin to glucose fluxes in the awake and unrestrained mouse. Metabolism **48**, 965–970 (1999).

26. Haluzik, M. M. et al. Improvement of insulin sensitivity after peroxisome proliferator-activated receptor-alpha agonist treatment is accompanied by paradoxical increase of circulating resistin levels. Endocrinology **147**, 4517–4524 (2006).

LETTERS

Organ regeneration does not require a functional stem cell niche in plants

Giovanni Sena[1], Xiaoning Wang[1], Hsiao-Yun Liu[1], Hugo Hofhuis[2] & Kenneth D. Birnbaum[1]

Plants rely on the maintenance of stem cell niches at their apices for the continuous growth of roots and shoots. However, although the developmental plasticity of plant cells has been demonstrated[1], it is not known whether the stem cell niche is required for organogenesis. Here we explore the capacity of a broad range of differentiating cells to regenerate an organ without the activity of a stem cell niche. Using a root-tip regeneration system in *Arabidopsis thaliana* to track the molecular and functional recovery of cell fates, we show that re-specification of lost cell identities begins within hours of excision and that the function of specialized cells is restored within one day. Critically, regeneration proceeds in plants with mutations that fail to maintain the stem cell niche. These results show that stem-cell-like properties that mediate complete organ regeneration are dispersed in plant meristems and are not restricted to niches, which nonetheless seem to be necessary for indeterminate growth. This regenerative reprogramming of an entire organ without transition to a stereotypical stem cell environment has intriguing parallels to recent reports of induced transdifferentiation of specific cell types in the adult organs of animals[2,3].

The indeterminate growth of plant organs arises from the activity of a localized stem cell niche, a micro-environment that supports stem cells[4,5]. In the plant root, longitudinal cell files converge on a stem cell niche comprised of a set of initials (stem cells) that are maintained in an undifferentiated state by contact with the quiescent centre, a group of cells with low mitotic activity (Fig. 1a). A newly formed quiescent centre is detected early after root-tip excision in pea and maize, and after quiescent centre laser ablation in *Arabidopsis*, which is consistent with the role of the niche as a pattern reorganizer in regeneration[6–8]. However, is the reconstitution of the stem cell niche the basis for the plant's high capacity to regenerate? Alternatively, can a wider population of cells have stem-cell-like properties, regenerating an organ independently of an actively dividing stem cell niche? Here we address the requirement for stem cell niche activity as a pattern organizer for organ regeneration.

To develop a comprehensive analysis of regeneration, we adapted root-tip excision techniques used in maize and pea[6,7] for use in *Arabidopsis*, enabling the examination of regeneration with high resolution using confocal imaging of cell-identity marker lines and well characterized mutants with meristematic defects. In combination, we used cell-type-specific transcriptional profiles generated previously[9–11] to track cell identities from microarray analysis of regenerating root tissue at specific time points after excision.

We performed standard excisions at 130 μm from the root tip, resulting in the complete removal of quiescent centre, all surrounding stem cells along with several tiers of daughter cells, and the root cap, including all of the columella and most of the lateral root cap (Fig. 1a; Methods). The standard excisions were made in a zone of proliferative cells that already express cell-specific markers[9]. No hormones or exogenous treatments were applied. Competence to regenerate extended to at least 200 μm from the root tip, with the frequency of regeneration dropping sharply at the proximal end of the meristematic zone, indicating an extended region of regeneration competence in the root tip (Fig. 1b).

Cell divisions during regeneration occurred in all major tissues constituting the root tip, as shown by analysis of a cell-cycle marker in five cell types or tissues (Fig. 1c and Supplementary Table 1, $n = 12$ roots). In addition, none of the fate-specific markers that we tracked by time-lapse imaging showed expanded expression patterns that could correlate with tissue-specific proliferation (Fig. 1d, e and Supplementary Figs 1 and 2). Cell division was required, because inhibition of the cell cycle prevented regeneration (Supplementary Fig. 3). However, re-patterning during regeneration did not seem to follow a stereotypical sequence of cell divisions, as in embryogenesis or lateral root formation. Taken together, these observations indicate that the meristematic zone as a whole, and not any specific tissue or cell type within it, contributes to root-tip regeneration.

To resolve the early timing of cell identity reappearance, we compared global transcriptional analysis of regenerating stumps with an existing library of cell-type-specific transcriptional profiles[9–11]. We sampled stumps for microarray analysis at 0 h, 5 h, 13 h, 22 h and 7 days after initial tip excision at 130 μm (Methods). Using cell-type-specific transcriptional analyses of the root, we identified sets of markers that were highly enriched in specific cell types, and analysed their activity during regeneration (Supplementary Table 2 and Methods). This technique permitted a highly sensitive measure of cell identity because early and late differentiation stage markers could be tracked using about 100 markers for each cell type (Fig. 1f). This global analysis of cell fate showed that molecular recovery of the excised cell identities had begun within five hours after cutting (Fig. 1f). For columella, the percentage recovery of enriched markers increased steadily compared to the stump at 0 h, reaching 21% at 5 h, 32% at 13 h, and 55% at 22 h (false discovery rate, $q < 5\%$, Methods), with demonstrated columella differentiation regulators, such as *AUXIN RESPONSE FACTOR 10*, induced at these early stages[12]. About 22% of quiescent centre identity recovered by 5 h ($q < 5\%$) without any further increase at 13 h and 22 h ($q < 5\%$). Thus, we can track the ordered re-establishment of cell identity, which shows the rapid re-specification of lost cell fates and identifies new candidate regulators for specification of cell identity (Supplementary Table 2). These results do not rule out that some quiescent-centre-specific genes may have a critical role in early regeneration, but they raise the question of whether differentiated cell types can be restored before the stem cell niche becomes functional.

We established the precise timing of the functional recovery of a completely excised cell type by focusing on columella cells, which reside at the tip of the root. In intact roots, differentiated columella cells accumulate starch within amyloplast organelles, a process required

[1]New York University, Department of Biology, Center for Genomics and Systems Biology, Silver Building, 100 Washington Square East, New York, New York 10003, USA. [2]Utrecht University, Department of Biology, Faculty of Science, Padualaan 8, 3584 CH, Utrecht, The Netherlands.
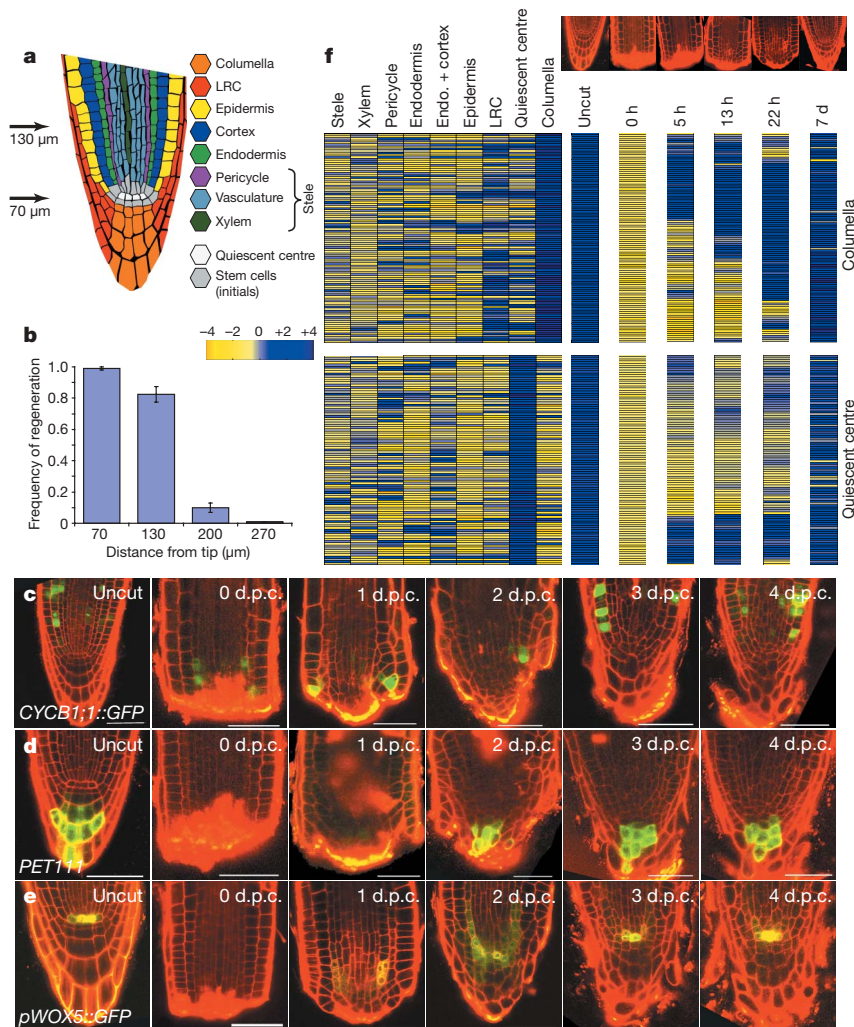
Figure 1 | Root-tip regeneration and cell fate re-specification in wild type. a, Schematic of *Arabidopsis* root apical meristem with quiescent centre (70 μm) and standard excision point (130 μm) positions; LRC, lateral root cap. **b**, Regeneration frequency in wild type (Col-0); $n = 102$ (70 μm), 57 (130 μm), 111 (200 μm), 32 (270 μm); error bars, standard error of the estimate of the proportion (Methods). **c–e**, Confocal time-lapse of single regenerating roots in *CYCB1;1::GFP* (**c**), the columella marker *PET111* (**d**) and the quiescent-centre-specific promoter fusion *pWOX5::GFP(ER)* (**e**), at consecutive days post cut (d.p.c.); scale bars, 50 μm. **f**, Expression levels of columella-enriched ($n = 103$, top panels) and quiescent-centre-enriched ($n = 95$, bottom panels) transcripts during regeneration; left, expression in cell types of uncut roots[9–11]; right, expression in uncut tips or regenerating stumps at the time points indicated.

for root gravitropism[13,14]. By one day post cut, Lugol staining confirmed *de novo* starch accumulation above the cut site (Fig. 2a and Supplementary Fig. 4). More intense staining was observed two days post cut (Fig. 2a and Supplementary Fig. 4). To test for recovery of columella function, we subjected regenerating roots to a standard gravitropism assay by reorienting them perpendicularly to the gravity vector and scoring the response over time. All wild-type roots showed a clear gravitropic response within 12 h. Although cut roots did not respond to gravity in the first 12 h after excision when cut at 130 μm, 13.8% of the cut roots exhibited a clear gravitropic response at 1 day post cut, 55.4% at 2 days post cut and 89.2% at 3 days post cut ($n = 65$, for all time points). However, the quiescent-centre-specific marker *WUSCHEL RELATED HOMEOBOX 5* (*WOX5*) was either ectopically expressed in the endodermal file or, at times, expressed in differentiated columella cells at one day post cut (Fig. 2a and Supplementary Fig. 4). Thus, as early as one day after complete columella excision, a new set of cells expressed columella markers and performed columella-specific functions while the morphology of the stem cell niche had not yet recovered.

Given the early re-establishment of a differentiated cell type, we tested the requirement for functional stem cells by using mutants in which post-embryonic root growth ceases due to the failure to maintain the stem cell niche. The *PLETHORA* (*PLT*) gene family has been shown to be critical for root formation[15], with the double mutant *plt1plt2* showing differentiation of stem cells at three days post germination[5], as verified under our conditions (Fig. 2b, note the lack of stem cell layer between the quiescent centre and the starch-stained columella). The uncut double mutant root has abnormal tip and stem cell niche morphology but normal gravitropism and convergent longitudinal cell files[5]. Surprisingly, *plt1plt2* roots cut at four days post

germination quickly regenerated by re-establishing the U-shaped convergent pattern of longitudinal cell files at the tip (Fig. 2c, e and Supplementary Fig. 5a). Moreover, starch granules accumulated in the regenerating double mutants (Fig. 2d) and the gravitropic response was re-established (Fig. 2c and Supplementary Fig. 6), indicating that functional columella cells were re-specified during regeneration. Similarly, *scarecrow* (*scr*) mutants, which fail to maintain root stem cell function through a pathway independent of *PLT1* and *PLT2* (refs 5, 16, 17), were also able to restore their pre-cut pattern, starch staining and gravitropism (Fig. 2b–d, f and Supplementary Figs 5b and 6). *PLT1* and *PLT2* are expressed early in regeneration in wild-type roots (Supplementary Fig. 7). However, using microarray comparison of *plt1plt2* mutant and wild-type roots, we ruled out that *PLT1*- and *PLT2*-dependent genes were induced by alternative mechanisms in regenerating double mutants (Supplementary Fig. 8). We note that a lower percentage of *plt1plt2* and *scr* mutants regenerated compared to wild-type roots (Fig. 2c), which we hypothesize is due to the documented effect of both mutants in reducing cell divisions in the meristematic zone[15,17]—the pool of cells recruited for regeneration. Together, these results show that stem cell niche activity is not necessary for early root-tip regeneration and imply the existence of an independent mechanism for cell specification and patterning in the meristematic region.

Several results suggest that auxin, which has been shown to position the root stem cell niche and to form a potentially instructive concentration gradient[18,19], may be a critical component of the mechanism that coordinates organogenesis[20]. First, roots failed to regenerate beyond the earliest stages when we blocked auxin transport during regeneration using N-1-naphthylphthalamic acid (NPA; Supplementary Fig. 9). Second, auxin efflux carriers and an auxin-responsive reporter
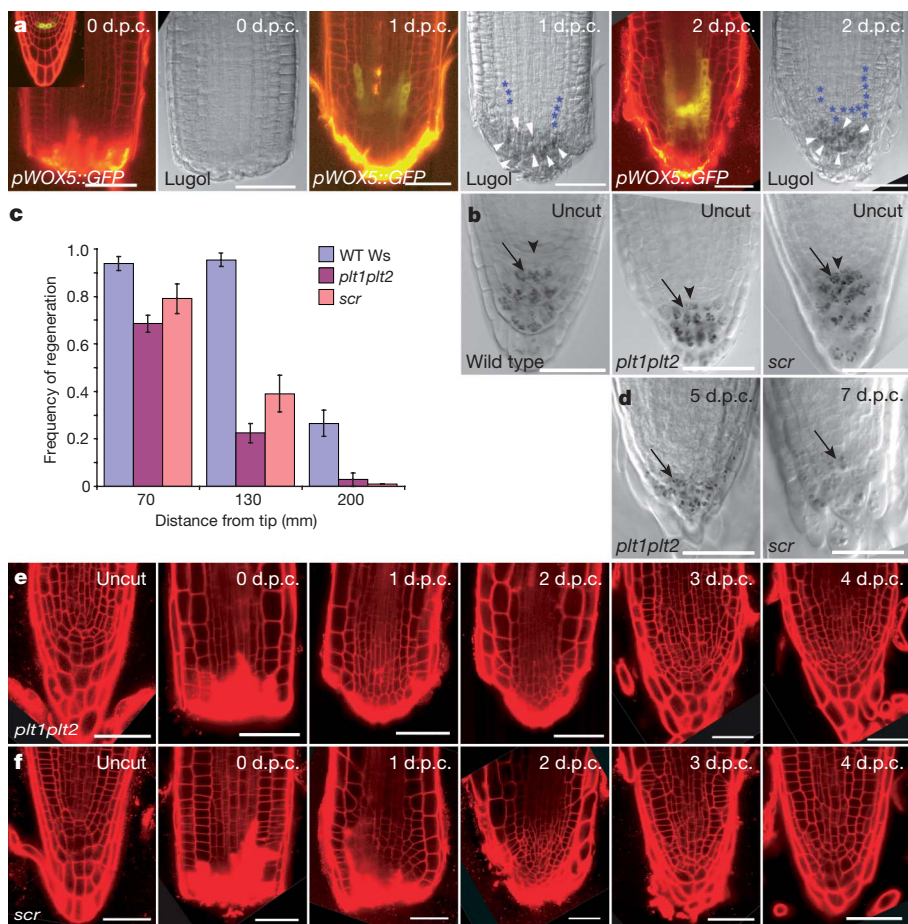
Figure 2 | Columella starch staining and root-tip regeneration in stem cell mutants.
a, Simultaneous confocal/differential interference contrast imaging of regenerating roots expressing *pWOX5::GFP(ER)*; inset, uncut root tip; blue asterisks indicate GFP-expressing cells as shown in the corresponding confocal image; white arrowheads indicate starch-stained cells; scale bars, 50 μm. **b**, **d**, Lugol staining of uncut (**b**) and regenerating (**d**) roots in wild type, *plt1plt2* and *scr*. Arrowheads indicate the position of quiescent centre; arrows indicate starch-stained cells; scale bars, 50 μm. **c**, Frequency of regeneration in wild type (WT Ws), *scr* and *plt1plt2* mutants; wild type, *n* = 66, 65 and 64 at 70 μm, 130 μm and 200 μm, respectively; *scr*, *n* = 43, 41 and 23; *plt1plt2*, *n* = 162, 107 and 35; error bars, standard error of the estimate of the proportion. **e**, **f**, Confocal time-lapse of regenerating roots in *plt1plt2* (**e**) and *scr* (**f**); scale bars, 50 μm.

re-established their excised domains at the root tip within a day of their excision (Fig. 3a–d). Third, many but not all genes induced in the first 24 h after excision have been shown previously to respond to auxin (Supplementary Table 2).

If organ regeneration does not require the activity of a stem cell niche, we hypothesized that other determinate organs should be capable of regeneration after excision. We developed a set of markers to distinguish competent versus non-competent tissue using transcriptional data on root developmental zones and a time-course induction of pluripotent callus from mature tissue[21] (Methods). Intriguingly, many of these markers showed high expression in young *Arabidopsis* leaves (9 days), compared to older leaves (15 and 22 days)[22] (Fig. 4a), indicating that young but not old leaves may be competent to regenerate, as suggested by historic reports[23]. Consistent with this prediction, we observed leaf regeneration in *Arabidopsis* after excising half of the leaf perpendicular to its midvein, in leaves corresponding to young stages (33.3%, *n* = 27) but never in leaves corresponding to older stages (*n* = 10, Fig. 4b, c). These observations suggest that the competence to re-pattern complex tissues may be a feature of many differentiating plant cells that share a common set of molecular properties.

What distinguishes these regeneration-competent cells from the stereotypical stem cells of the niche? In the *Arabidopsis* root, a body of work has shown that the stem cell niche is critical for indeterminate growth[5,8,17], which was not restored during regeneration in the *plt1plt2* and *scr* mutants. This indicates that continuous growth may be a unique feature of the stem cell niche whereas organogenesis is not.

The convergence of organ patterning and growth at the stem cell niche of *Arabidopsis* has made it difficult to separate these two

Figure 3 | Early auxin distribution in the regenerating root tip.
a–d, Confocal time-lapse of single regenerating roots expressing the auxin-responsive reporter *DR5::GFP(ER)* (**a**) and translational fusions of the auxin efflux carriers *pPIN1::PIN1::GFP* (**b**), *pPIN2::PIN2::GFP* (**c**) and *pPIN7::PIN7::GFP* (**d**) at consecutive days post cut. Arrowheads in **d** indicate distal expression domains; scale bars, 50 μm.
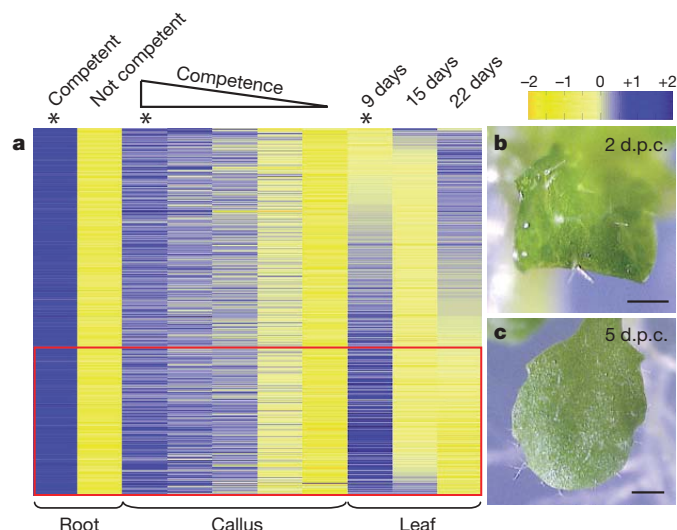
1152

Figure 4 | Regeneration competence markers and leaf regeneration.
a, Expression levels of regeneration-competence markers identified by enrichment in competent tissue in root and callus[21] (0-, 2-, 4-, 7- and 10-day-old tissue treated with auxin, $n = 647$, see Methods), showing high expression of competence markers in young but not in older leaf tissue. The red box indicates genes in the root/callus-competent marker set that also matched a putative competence profile for young leaves (209 genes, Supplementary Table 1 and Methods)[22]. Asterisks indicate states of highest competence. **b, c,** Regenerating single four-day-old leaf, at two (**b**) and five (**c**) days post cut; scale bars, 50 μm.

fundamental processes. Taken together, our results separate a widely dispersed capacity for pluripotency and patterning during organogenesis from the narrowly located capacity for indeterminate growth within the stem cell niche. The extension of stem-cell-like properties that mediate organogenesis in maturing tissues may predispose the plant for a high capacity to regenerate. Recent work has shown that adult mammalian cells may also be induced to directly switch fates without stem cell intermediates[2,3]. Plants and perhaps other highly regenerative organisms seem to be able to reprogram entire organs in this way. These findings provide a new basis to search for mechanisms that coordinate organogenesis independently of a central organizer.

METHODS SUMMARY

Mutant alleles used were *plt1-4 plt2-2* in the ecotype Wassilewskija (Ws) and *scr-4* (Ws). Seedlings at four days post germination were excised by hand under a dissecting microscope using a 30G sterile dental needle (ExelInt). The frequency of regeneration was defined as the fraction of the plants that showed root regeneration at six days post cut, measured by gravitropic response and confirmed by tip morphology. For microarray analysis, 130 μm of root tips were removed to instigate regeneration and then, during tip regeneration, 70 μm of regenerating stumps were manually dissected at the indicated time points.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Steward, F. C., Ammirato, P. V. & Mapes, M. O. Growth and development of totipotent cells — some problems, procedures, perspectives. *Ann. Bot. (Lond.)* **34**, 761–787 (1970).
2. Thowfeequ, S., Myatt, E. J. & Tosh, D. Transdifferentiation in developmental biology, disease, and in therapy. *Dev. Dyn.* **236**, 3208–3217 (2007).
3. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. *In vivo* reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* **455**, 627–632 (2008).
4. van den Berg, C., Willemsen, V., Hendriks, G., Weisbeek, P. & Scheres, B. Short-range control of cell differentiation in the *Arabidopsis* root meristem. *Nature* **390**, 287–289 (1997).
5. Aida, M. *et al.* The PLETHORA genes mediate patterning of the *Arabidopsis* root stem cell niche. *Cell* **119**, 109–120 (2004).
6. Feldman, L. J. *De novo* origin of quiescent center regenerating root apices of *Zea mays*. *Planta* **128**, 207–212 (1976).
7. Rost, T. L. & Jones, T. J. Pea root regeneration after tip excisions at different levels — polarity of new growth. *Ann. Bot. (Lond.)* **61**, 513–523 (1988).
8. Xu, J. *et al.* A molecular framework for plant regeneration. *Science* **311**, 385–388 (2006).
9. Birnbaum, K. *et al.* A gene expression map of the *Arabidopsis* root. *Science* **302**, 1956–1960 (2003).
10. Nawy, T. *et al.* Transcriptional profile of the *Arabidopsis* root quiescent center. *Plant Cell* **17**, 1908–1925 (2005).
11. Lee, J. Y. *et al.* Transcriptional and posttranscriptional regulation of transcription factor expression in *Arabidopsis* roots. *Proc. Natl Acad. Sci. USA* **103**, 6055–6060 (2006).
12. Wang, J. W. *et al.* Control of root cap formation by microRNA-targeted auxin response factors in *Arabidopsis*. *Plant Cell* **17**, 2204–2216 (2005).
13. Sack, F. D. Plastids and gravitropic sensing. *Planta* **203**, S63–S68 (1997).
14. Blancaflor, E. B., Fasano, J. M. & Gilroy, S. Mapping the functional roles of cap cells in the response of *Arabidopsis* primary roots to gravity. *Plant Physiol.* **116**, 213–222 (1998).
15. Galinha, C. *et al.* PLETHORA proteins as dose-dependent master regulators of *Arabidopsis* root development. *Nature* **449**, 1053–1057 (2007).
16. Heidstra, R., Welch, D. & Scheres, B. Mosaic analyses using marked activation and deletion clones dissect *Arabidopsis* SCARECROW action in asymmetric cell division. *Genes Dev.* **18**, 1964–1969 (2004).
17. Sabatini, S., Heidstra, R., Wildwater, M. & Scheres, B. SCARECROW is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. *Genes Dev.* **17**, 354–358 (2003).
18. Sabatini, S. *et al.* An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root. *Cell* **99**, 463–472 (1999).
19. Grieneisen, V. A., Xu, J., Maree, A. F., Hogeweg, P. & Scheres, B. Auxin transport is sufficient to generate a maximum and gradient guiding root growth. *Nature* **449**, 1008–1013 (2007).
20. Benkova, E. *et al.* Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell* **115**, 591–602 (2003).
21. Che, P., Lall, S., Nettleton, D. & Howell, S. H. Gene expression programs during shoot, root, and callus development in *Arabidopsis* tissue culture. *Plant Physiol.* **141**, 620–637 (2006).
22. Beemster, G. T., Vercruysse, S., De Veylder, L., Kuiper, M. & Inze, D. The *Arabidopsis* leaf as a model system for investigating the role of cell cycle regulation in organ growth. *J. Plant Res.* **119**, 43–50 (2006).
23. Figdor, W. Uber Restitutionserscheinungen an Blattern von Gesneriaceen. *Jahrb. Wiss. Bot.* **44**, 41–56 (1907).

Author Contributions G.S. performed all experiments with help from X.W. and H.Y.L. on dissections and microarrays. H.H. generated *PLT* reporter constructs. K.D.B. and G.S. conceived the project, planned all experiments, performed data analysis and wrote the manuscript.

Author Information Microarray data have been submitted to GEO (NCBI) under the accession number GSE9996. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.D.B. (ken.birnbaum@nyu.edu).

METHODS

Microarray and statistical analysis. Microarray profiles were normalized using the MAS 5.0 method with a target intensity of 250. Cell-type-specific marker sets were generated by identifying transcripts for which the signal was significantly enriched in a given cell type compared to all other cell types, using significance analysis of microarrays (SAM) with a false discovery rate (q) cutoff $<5\%$[24] and a twofold enrichment cutoff. To increase stringency for cell specificity and assure no overlap between columella and quiescent centre markers, we also required a twofold enrichment in columella average signal over the average signal in each of the other cell types of the root tip (for example, columella markers were twofold enriched over quiescent centre and lateral root cap, respectively). The same procedure was followed for quiescent centre markers, ensuring a twofold enrichment over columella and lateral root cap. In addition, the root-tip-specific cell types also needed to show a twofold enrichment in root tip over proximal meristem expression.

For analysis of percentage columella and quiescent centre identity recovery, ranked gene expression was tested for a significant fit to modelled expression patterns representing an increase in expression at either 5 h, 13 h or 22 h using the quantitative test in SAM ($q < 5\%$). For example, genes that increase significantly at the 5 h regeneration time point fit the pattern 1 2 2 2, where 1 represents expression of replicates at time 0 and 2 represents replicates at the subsequent time points of regeneration (5 h, 13 h and 22 h). The rank method in SAM was used.

For evaluating *PLT* downstream markers, a two-class unpaired test in SAM ($q < 5\%$) was used to find quiescent centre markers significantly downregulated in the *plt1plt2* mutant tips compared to wild-type tips (termed the *PLT1/2*-dependent set). Subsequent analysis tested whether any members of the *PLT1/2*-dependent set were significantly upregulated in wild-type stumps at 24 h versus wild-type stumps at 0 h (testing for early regulation of the *PLT1/2*-dependent set in wild type) or wild-type stumps at 24 h versus *plt1plt2* stumps at 0 h (testing for potential regulation of the *PLT1/2*-dependent set in the *plt1plt2* mutant during regeneration, that is, alternate regulatory mechanisms) using the two-class unpaired test in SAM ($q < 5\%$).

Lists of competence markers for root, callus and leaves were generated sequentially, and the intersection of each set was taken. To generate root-competence markers, a two-class unpaired test in SAM was performed to find genes significantly upregulated in tissue freshly collected at 130–200 µm (competent zone) versus tissue freshly collected at 270–340 µm (non-competent zone) with a $q < 5\%$. This procedure yielded 1,538 genes (root-competence markers). To identify competence markers in tissue explants undergoing auxin treatment to generate callus, a quantitative analysis in SAM ($q < 5\%$) was used querying for genes that showed a monotonic increase in the callus induction samples over days 0, 2, 4, 7 and 10 on callus-inducing media (CIM) with data from previous work[21] using the rank method so that replicates for each time point were labelled: 1, 2, 3, 4 and 5, respectively (callus-competent markers). The intersection of the root- and callus-competent sets was 647 genes. To identify potential competence markers in leaf, genes significantly upregulated in 9-day-old leaves versus 22-day-old leaves[22] were determined using a two-class unpaired test in SAM ($q < 5\%$; leaf-competent markers). The intersection of the root-, callus- and leaf-competent marker sets was 209 genes.

To find all genes that were significantly regulated in regenerating stumps five hours after tip-cutting, we used a two-class unpaired test in SAM ($q < 5\%$) comparing replicates in regenerating tips at 0 h versus 5 h. We found the intersection of that list and the list of auxin-induced genes[25] to generate the list of genes induced at 5 h after tip cutting that were also induced by auxin. Among the genes that were differentially regulated in root stumps in the first five hours after cutting ($n = 182$, Supplementary Table 2), 22 have been shown to respond to auxin[25].

The standard error of the estimate of the proportion is the standard deviation of the population of all possible values of the proportion computed from samples of a given size n. Given P, the estimate of the true proportion as calculated from one random sample of size n, the standard error of the estimate of the proportion is estimated as the square root of $P*(1 - P)/n$.

24. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
25. Nemhauser, J. L., Hong, F. & Chory, J. Different plant hormones regulate similar processes through largely nonoverlapping transcriptional responses. *Cell* **126**, 467–475 (2006).

LETTERS

Ca^{2+}/calmodulin regulates salicylic-acid-mediated plant immunity

Liqun Du[1], Gul S. Ali[3], Kayla A. Simons[1], Jingguo Hou[2]†, Tianbao Yang[1], A. S. N. Reddy[3] & B. W. Poovaiah[1]

Intracellular calcium transients during plant–pathogen interactions are necessary early events leading to local and systemic acquired resistance[1]. Salicylic acid, a critical messenger, is also required for both of these responses[2,3], but whether and how salicylic acid level is regulated by Ca^{2+} signalling during plant–pathogen interaction is unclear. Here we report a mechanism connecting Ca^{2+} signal to salicylic-acid-mediated immune response through calmodulin, AtSR1 (also known as CAMTA3), a Ca^{2+}/calmodulin-binding transcription factor, and EDS1, an established regulator of salicylic acid level. Constitutive disease resistance and elevated levels of salicylic acid in loss-of-function alleles of *Arabidopsis AtSR1* suggest that AtSR1 is a negative regulator of plant immunity. This was confirmed by epistasis analysis with mutants of compromised salicylic acid accumulation and disease resistance. We show that AtSR1 interacts with the promoter of *EDS1* and represses its expression. Furthermore, Ca^{2+}/calmodulin-binding to AtSR1 is required for suppression of plant defence, indicating a direct role for Ca^{2+}/calmodulin in regulating the function of AtSR1. These results reveal a previously unknown regulatory mechanism linking Ca^{2+} signalling to salicylic acid level.

Arabidopsis thaliana signal responsive (*AtSR*) proteins belong to a class of Ca^{2+}/calmodulin (CaM)-binding transcription factors[4–7]. In animals, AtSR homologues are involved in diverse functions[8,9]. Although AtSR proteins are implicated in plant responses to stresses[6], the specific function of AtSRs remains unknown. To determine the function of AtSR1, we isolated two loss-of-function mutants (*Atsr1-1* and *Atsr1-2*) in *Arabidopsis* (Supplementary Figs 1a and 2a, e). At 25–27 °C (12 h or other photoperiods), no noticeable difference between the wild-type and *Atsr1* mutants was observed (Fig. 1a). However, *Atsr1-1* showed elevated resistance to virulent *Pseudomonas syringae* pv. *tomato* DC3000 (hereafter called *Pst* DC3000; Fig. 1b), as well as avirulent *Pst* (*AvrRpt2*; data not shown). Because elevated resistance to pathogens is usually correlated with induced expression of *PR* genes[10], we analysed the expression of *PR1* in wild-type and *Atsr1-1* plants inoculated with *Pst* DC3000. In wild-type plants *PR1* expression did not start until 24 h after inoculation, whereas in *Atsr1-1* mutants its expression started 6 h after inoculation. However, the maximum expression of *PR1* remained similar between wild-type and *Atsr1* plants (Fig. 1c). The elevated disease resistance and sensitized *PR1* expression in *Atsr1-1* plants indicate that it is a repressor of plant immunity.

At 19–21 °C (12-h photoperiod), *Atsr1-1* plants showed reduced growth (Fig. 1a, d and Supplementary Fig. 2d). The expression of systemic acquired resistance-associated marker genes[10], *PR1*, *PR2* and *PR5*, was constitutively activated under lower temperature in *Atsr1-1* plants (Fig. 1f). Predictably, the disease resistance of *Atsr1-1* plants was also enhanced in comparison with plants grown under higher temperature (Fig. 1b, e). The *Atsr1* plants displayed chlorosis

and autonomous lesions (Fig. 1g and Supplementary Figs 2b, d). Plants undergoing hypersensitive response produce reactive oxygen species (ROS) and autofluorescent compounds[11,12]. Staining for H$_2$O$_2$ revealed numerous brown patches on *Atsr1-1* leaves, which were comparable to wild-type plants infected with incompatible *Pst* (*AvrRpt2*) (Fig. 1h and Supplementary Fig. 3). *Atsr1-1* leaves also showed extensive autofluorescence (Fig. 1i). These results indicate that *Atsr1* plants grown at a lower temperature show hallmarks of constitutive defence responses commonly found in lesion-mimicking mutants[11] or wild-type plants inoculated with avirulent bacterial pathogens[12]. The temperature-dependent autoimmunity further suggests that AtSR1 represses R-protein-mediated defence activation. Recent reports show that the stability of active R proteins is regulated by co-chaperone RAR1 in plants in a temperature-dependent manner, with lower temperatures favouring the accumulation of R proteins[13,14]. Conceivably, lower temperatures might favour the accumulation of some R proteins in *Arabidopsis*, but AtSR1 represses the mis-activation of defence whereas its absence in *Atsr1* does not. The expression of *AtSR1* cDNA in *Atsr1* mutants restored all mutant phenotypes (Supplementary Fig. 4), confirming that the *Atsr1* phenotypes are caused by loss of AtSR1.

Because *Atsr1* mutants resemble mutants with increased salicylic acid levels[11,15], we quantified salicylic acid in the mutant and wild-type plants grown at 19–21 °C. Free and conjugated salicylic acid levels were increased ∼7- and ∼8-fold, respectively, in the *Atsr1-1* plants (Fig. 2a, b). In uninfected plants grown at 25–27 °C, salicylic acid levels in *Atsr1* and wild-type plants were similar. However, the salicylic acid level increased faster in *Atsr1* than in wild-type plants when inoculated with *Pst* DC3000 (Supplementary Fig. 5). Previous studies have shown that elevating salicylic acid levels alone is enough to cause an enhanced immune response and reduced growth[2,16]. Expressing the salicylic-acid-degrading enzyme NahG suppressed both disease resistance and retarded growth in some disease-resistant mutants (*acd6*, *bon1* and *ssi1*) with elevated salicylic acid levels[15,17], but only disease resistance in other mutants (*mpk4* and *dnd1*)[18,19]. To determine if the reduced growth and enhanced disease resistance of *Atsr1* mutants are caused by elevated salicylic acid levels or other mechanism(s), we eliminated salicylic acid by expressing *NahG* in wild-type and *Atsr1-1* plants. Wild-type and *Atsr1-1* plants expressing *NahG* appeared to be similar but both were bigger than the wild type (Fig. 2c–e). Furthermore, constitutively activated *PR1* expression was blocked in *Atsr1-1 NahG* plants (Fig. 2c); both *Atsr1-1 NahG* and wild-type *NahG* plants were more sensitive to *Pst* DC3000 than wild-type plants (Fig. 2f). These results indicate that the elevated salicylic acid level is the major cause of *Atsr1-1* phenotypes.

Consistent with elevated salicylic acid levels, the expression of *ICS1*, *PAD4*, *EDS1* and *EDS5*, important positive regulators of salicylic acid biosynthesis, is highly induced in *Atsr1-1* plants

[1]Center for Integrated Biotechnology and Department of Horticulture, [2]Department of Chemistry, Washington State University, Pullman, Washington 99164-6414, USA. [3]Department of Biology and Program in Molecular Plant Biology, Colorado State University, Fort Collins, Colorado 80523-1878, USA. †Present address: Bioanalytical Services, Primera Analytical Solutions Corp., 259 Wall Street, Princeton, New Jersey 08540, USA.
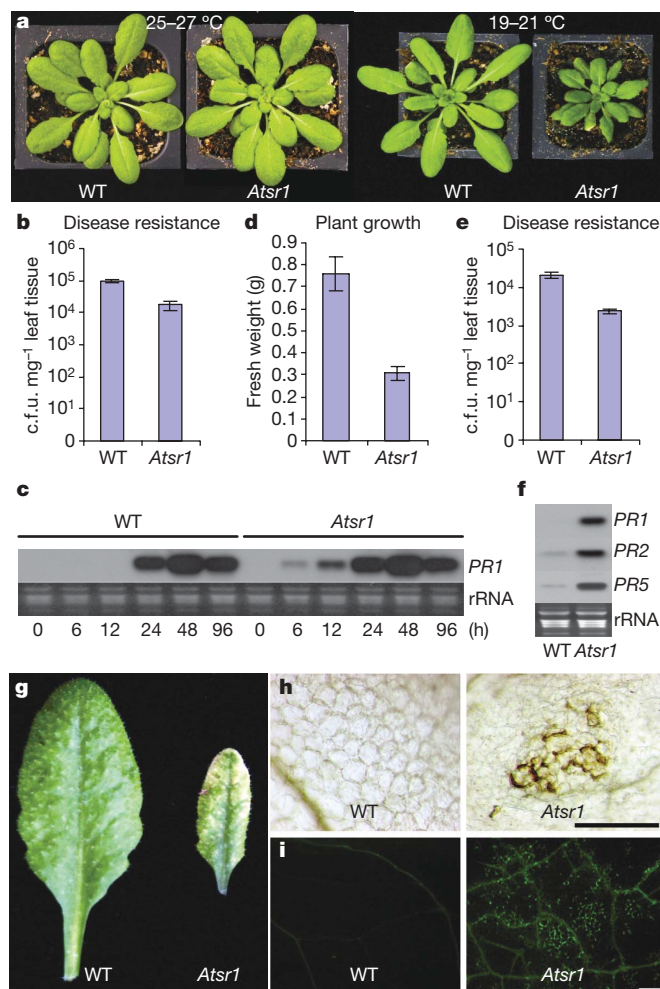
Figure 1 | The *Atsr1-1* mutant shows sensitized defence responses.
a, Phenotypic comparison. Thirty-two-day-old (left) and thirty-five-day-old (right) plants grown at indicated temperature. **b**, Disease resistance test. *Pst* DC3000 (optical density at 600 nm (OD_{600}) of 0.001) was infiltrated into rosette leaves (4 weeks, 25–27 °C), and the colony-forming units (c.f.u.) at 3 days post inoculation (d.p.i.) are shown. **c**, Time course of *PR1* expression. Rosette leaves (4-week-old plants, 25–27 °C) were infiltrated with *Pst* DC3000 (OD_{600} 0.001) and samples were taken at the indicated time for northern blot analysis. **d**, Growth comparison in terms of fresh rosette weight (5 weeks, 19–21 °C). **e**, Disease resistance test. *Pst* DC3000 (OD_{600} 0.0001) was infiltrated into rosette leaves (19–21 °C) and the c.f.u. at 3 d.p.i. are shown. **f**, Constitutive induction of *PR* genes in the *Atsr1-1* mutant. Total RNA samples were prepared from rosette leaves grown at 19–21 °C; identical blots were hybridized to indicated probes. **g**, Early chlorosis of *Atsr1-1* leaf compared with wild-type leaf of the same age. **h**, Production of H_2O_2 (brown patches). 3,3′-diaminobenzidine (DAB) was used to stain leaves of a similar age from wild-type (left) and *Atsr1-1* (right) plants. Original magnification, 10 × 40. **i**, Autofluorescence image of wild type (left) and *Atsr1-1* (right). Original magnification, 10 × 10. All plants were grown under 12-h photoperiod, all data are expressed as mean ± s.d. ($n = 4$); ethidium-bromide-stained rRNA was used as loading control for all northern blots (rRNA). Scale bars, 200 μm.

Figure 2 | The pleiotropic phenotype of *Atsr1-1* is dependent on salicylic acid. **a**, **b**, Free (**a**) and conjugated (**b**) salicylic acid (SA) quantification. Rosette leaves from 5-week-old plants were used for salicylic acid quantification. **c**, Molecular characterization demonstrating the genotype of the tested plant lines. Genotype is marked beneath the panels. For definitions of PCR1 and PCR2, see the legend of Supplementary Fig. 2a. *NahG* and *PR1* expression is shown in northern blots; rRNA is used as loading control. **d**, Phenotypic comparison of wild type, *Atsr1-1*, wild type transformed with *35S::NahG* (WT *NahG*) and *Atsr1-1* transformed with *35S::NahG* (*Atsr1 NahG*). **e**, Growth comparison in terms of fresh rosette weight. **f**, Disease resistance test. *Pst* DC3000 (OD_{600} 0.0001) was infiltrated into rosette leaves and the c.f.u. at 3 d.p.i. is shown. All plants were grown at 19–21 °C with 12-h photoperiod for 5 weeks, and all data are expressed as mean ± s.d. ($n = 4$).

Arabidopsis ecotype Columbia[17], it was not used for epistasis analysis. The *pad4* and *ics1* single mutants, as well as the *Atsr1 pad4* and *Atsr1 ics1* double mutants, were more sensitive to *Pst* DC3000 than the wild type (Supplementary Fig. 6c). Furthermore, in the double mutants, constitutive expression of *PR1* and the dwarf phenotype of *Atsr1-1* are restored to wild-type levels (Supplementary Fig. 6d, e). The *eds5* mutation also blocked the *Atsr1* phenotypes (data not shown). These results suggest that *AtSR1* functions at a step no later than *PAD4* in the salicylic acid activation cascade.

AtSR1 and its homologues bind to the conserved CGCG box and regulate the expression of target genes[6,8]. Analysis of *ICS1*, *PAD4*, *EDS1* and *EDS5* promoters revealed a typical CGCG box (ACGCGT) only in the *EDS1* promoter (−746 to −741, Supplementary Fig. 7a), indicating a direct regulation of *EDS1* by AtSR1. We showed that the AtSR1 DNA-binding domain (1–153 amino acids) binds the *EDS1* promoter fragment (−762 to −731) in an ACGCGT-dependent (Fig. 3a) and Ca^{2+}/CaM-independent manner (data not shown). Chromatin immunoprecipitation assay further confirmed that a full-length AtSR1–YFP interacts with the *EDS1* promoter *in vivo* (Fig. 3a). To study the functional significance of AtSR1-binding to the *EDS1* promoter, the −1.5-kb promoter of *EDS1* (*EDS1P*) was cloned and the ACGCGT element was mutated to ACCCGT (*eds1p*) to abolish its interaction with AtSR1. Both *EDS1P* and *eds1p* were fused to *luciferase* (*luc*) and expressed in protoplasts of

(Supplementary Fig. 6a). *ICS1*, *PAD4*, *EDS1* and *EDS5* are arranged in a sequential order of PAD4/EDS1, EDS5 and ICS1 (refs 2, 20, 21). The expression of *EDS1*, *PAD4* and *EDS5* is induced by salicylic acid[2] and *ICS1* is induced in the constitutively resistant mutant[22]. Therefore, the elevated expression of these genes in the *Atsr1* mutant could either be the direct result of a failure in AtSR1 regulation, or merely the consequence of constitutively activated immunity and elevated salicylic acid levels. Epistasis analysis between *Atsr1* and these regulators could lead us closer to the AtSR1-regulated step. Because there are two closely linked functional *EDS1* genes in the
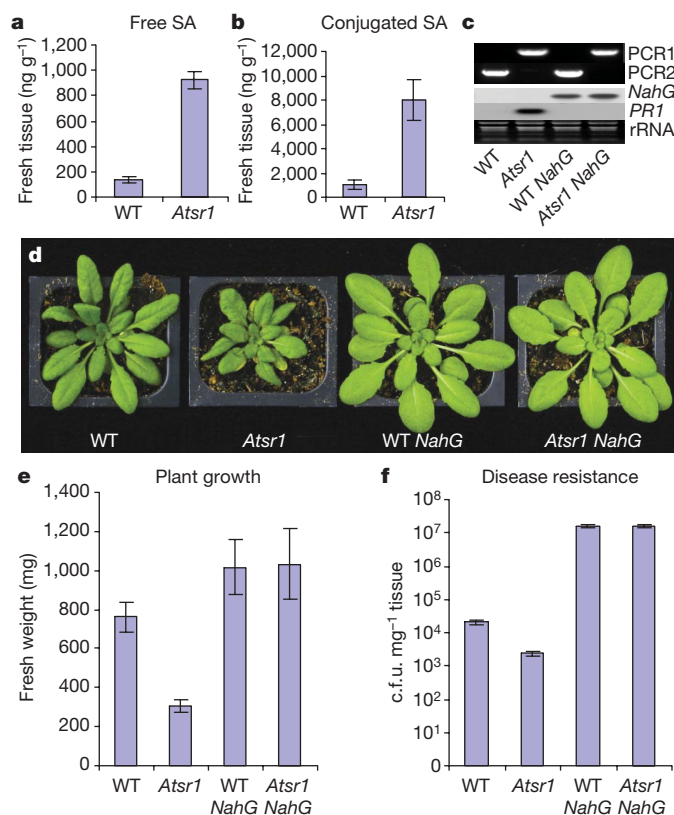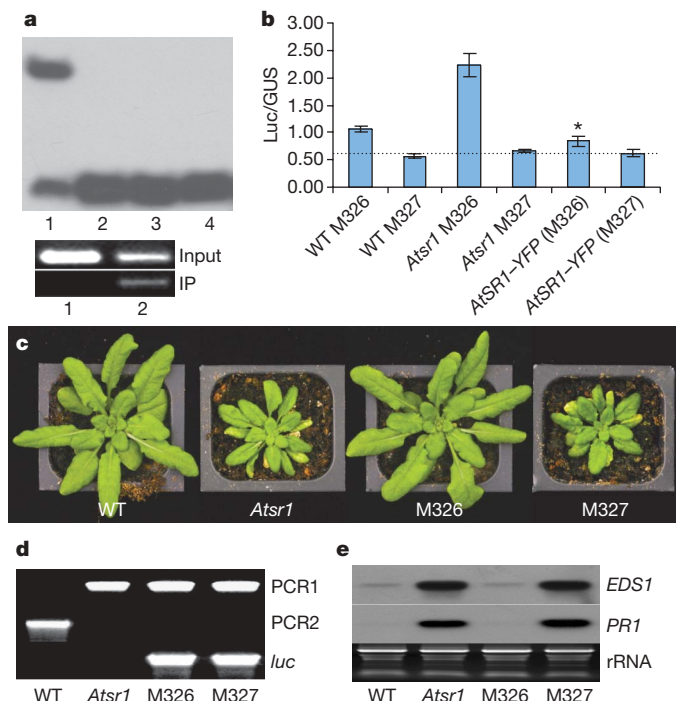
1155

Figure 3 | AtSR1 is involved in transcriptional regulation of EDS1. a, AtSR1 binds the *EDS1* promoter. Top panel: the *EDS1* promoter fragment (*EDS1P*, −762 to −731) or a mutated version with its ACGCGT changed to ACCCGT (*eds1p*) was used in an EMSA. Lane 1, ^{32}P-labelled *EDS1P* plus AtSR1; lane 2, ^{32}P-labelled *EDS1P* plus 500-fold unlabelled *EDS1P* plus AtSR1; lane 3, ^{32}P-labelled *eds1p* plus AtSR1; lane 4, ^{32}P-labelled *EDS1P* only. Bottom panel: chromatin immunoprecipitation (IP) analysis. The chromatin immunoprecipitated DNA was amplified with primers flanking the CGCG box in the *EDS1* promoter. Lane 1, wild-type protoplasts expressing YFP; lane 2, wild-type protoplasts expressing AtSR1–YFP. **b**, Promoter activity assay of wild-type (M326, *EDS1P::luc*) and mutated (M327, *eds1p::luc*) *EDS1* promoter constructs in protoplasts of wild-type, *Atsr1-1* and *Atsr1-1* expressing AtSR1–YFP plants. The ratio of Luc/GUS activities was measured. Data are expressed as mean ± s.d. (*n* = 4, asterisk indicates *P* < 0.039 by *t*-test). **c**, Phenotypic comparison of 39-day-old plants; genotype is marked beneath the panel. M326, *Atsr1-1* carrying *EDS1P::luc*; M327, *Atsr1-1* carrying *eds1p::luc*. **d**, Genotyping of plants shown in **c**. See the legend of Supplementary Fig. 2a for definitions of PCR1 and PCR2. Luc, luciferase band amplified with Luc-I (574–598 bp) and Luc-R (1,653–1,625 bp) primers. **e**, *EDS1* and *PR1* expression revealed by northern blot; rRNA was used as a loading control.

wild type, *Atsr1-1* and *Atsr1-1* expressing 35S::AtSR1–YFP. The *EDS1P::luc* (pDL326) activity was about twofold higher in *Atsr1-1* than in wild-type plants (Fig. 3b), indicating that AtSR1 negatively regulates *EDS1*. Predictably, *EDS1* promoter activity in *Atsr1-1* protoplasts overexpressing AtSR1 is reduced to a level slightly lower than that in wild type (Fig. 3b). This, together with a recent report that elevated expression of *EDS1* alone is adequate to constitutively activate immunity[23], indicate that the derepression of *EDS1* in *Atsr1* plants could have caused the constitutive immunity. Logically, *eds1p* does not bind AtSR1 and should result in elevated activity even in wild-type plants if AtSR1 is the only *trans*-acting factor binding to the mutated region. Notably, the activity of *eds1p::luc* (pDL327) is significantly lower than that of *EDS1P* and decreased to similar levels in all three cases (Fig. 3b). These results suggest that the ACCCGT mutation may have interfered with the binding of other *trans*-acting factor(s) to the CGCG box or a recognition core overlapping it, which is essential for the basal and/or induced transcription of *EDS1*. Our data suggest that these kinds of transcription factors do exist in *Arabidopsis* (Supplementary Fig. 7b). A model presented in Supplementary Fig. 7c illustrates the regulation of *EDS1* promoter.

The activity of *EDS1::luc* is ~2-fold higher in *Atsr1* plants in comparison with wild-type plants (Fig. 3b), noticeably less than the

four- to fivefold difference revealed by northern blot analysis (Supplementary Fig. 6a). In addition to the diluted feedback induction of *EDS1* by salicylic acid or other messengers in protoplast maintaining buffer, the introduction of an extra copy (or copies) of *EDS1* promoter into *Atsr1* mutants may have reduced the induction of *EDS1P::luc* as well as endogenous *EDS1*, because the elevated expression of *EDS1* in *Atsr1* mutants is driven by unidentified positive transcription factor(s) (see model, Supplementary Fig. 7c). To test this, we generated stable wild-type and *Atsr1* transformants carrying pDL326 or pDL327. All wild-type plants carrying pDL326 or pDL327 grew like wild-type plants without these plasmids (data not shown); all the *Atsr1-1* plants carrying pDL327 (M327) grew like *Atsr1-1* plants without the plasmid (Fig. 3c and Supplementary Fig. 8b). Notably, most of the *Atsr1-1* plants carrying pDL326 (M326) showed varying degrees of phenotypic rescue (Supplementary Fig. 8a). Nearly 10% of them grew like wild-type plants during their entire life cycle (Fig. 3c and Supplementary Fig. 8a) and lacked *AtSR1* (Fig. 3d). Remarkably, the expression of endogenous *EDS1* in these lines was restored to wild-type levels (Fig. 3e and Supplementary Fig. 8c), indicating that the phenotypic restoration is due to the quenched *EDS1* expression. Consistently, constitutive *PR1* expression was also abolished in rescued M326 lines (Fig. 3e and Supplementary Fig. 8c). Segregation analysis of T2 progeny of M326 lines indicated that the phenotypic rescue is mostly correlated with particular insertion events rather than a dosage effect of insertion (data not shown). It seems that insertion of *EDS1* promoter at some particular positions in the *Atsr1* genome competes for the ACGCGT-binding positive regulator(s) and quenches the endogenous *EDS1* expression, although the precise mechanism remains to be resolved. Failure of *eds1p::luc* to rescue the *Atsr1* phenotype (Fig. 3c, e and Supplementary Fig. 8b) further supports this notion.

Functional tests of mutations in a null mutant background[24,25] provide an effective strategy to study regulation of AtSR1 function by Ca[2+]/CaM. Three mutations (M1, I909V; M2, K907E; M3, Δ900–922) in the calmodulin-binding domain (CaMBD) of AtSR1 (refs 5–7) were generated (Supplementary Fig. 1b). Wild-type AtSR1 and AtSR1(I909V) bound CaM, whereas AtSR1(K907E) and AtSR1(Δ900–922) did not (Fig. 4a). Wild-type and mutated *Atsr1* plants were fused to Flag tag (Supplementary Fig. 1b), and expressed in *Atsr1-1* mutants. Most of the T1 plants (>30) complemented with 35S::*Atsr1(I909V)* (cM1) showed a rescued phenotype. None of the >30 individual T1 plants complemented with either 35S::*Atsr1(K907E)* (cM2) or 35S::*Atsr1(Δ900–922)* (cM3) was restored to the wild-type growth level. For accurate comparison of all the complemented lines, T2 plants with verified genotype and similar transgene expression (Fig. 4b) were compared for their phenotypes. The *Atsr1-1* mutants complemented with 35S::*AtSR1* (cW) or cM1 plants were restored to wild type with regards to their morphology (Fig. 4c), growth (Fig. 4d) and disease resistance (Fig. 4e). The level of salicylic acid in cW and cM1 plants was ~60% of wild type (Fig. 4f), and expression of *PR* and salicylic acid signalling genes in cW and cM1 plants was also slightly lower than in wild type (Fig. 4g). However, cM2 and cM3 plants resembled the *Atsr1-1* plants with chlorosis (Fig. 4c) and slightly increased growth (Fig. 4d). The level of salicylic acid in cM2 and cM3 plants was slightly lower than that in *Atsr1-1* plants, but still markedly higher than that in wild-type plants (Fig. 4f). The level of disease resistance of cM2 and cM3 plants was similar to that of *Atsr1-1* plants (Fig. 4e). Expression of *PR* and salicylic acid signalling genes in cM2 and cM3 plants was also slightly lower than in *Atsr1* plants but significantly higher than in wild-type plants (Fig. 4g). The fact that *Atsr1* mutants that lost their CaM-binding activity are compromised in their function suggests that Ca[2+]/CaM-binding is required for AtSR1 to suppress plant immunity.

Plant immunity is a well-balanced process in which Ca[2+]-mediated signalling is actively involved. It is documented that Ca[2+]-mediated signalling stimulates the production of ROS and NO, which positively regulate hypersensitive response and plant defence[1,26,27]. Positive regulators in the plant defence machinery are amplified in multiple
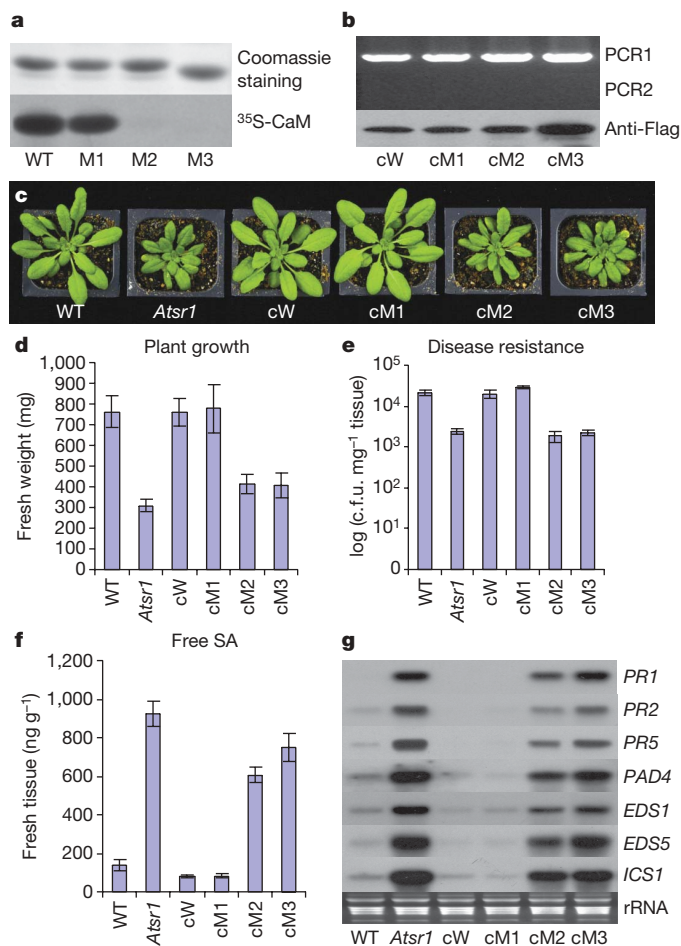
Figure 4 | Repression of immune response by AtSR1 is regulated by Ca²⁺/CaM. a, ³⁵S-labelled CaM binding assay. Purified recombinant proteins of wild type (WT) and three mutated versions of AtSR1 (I909V (M1), K907E (M2) and Δ900–922 (M3)) were bound to ³⁵S-labelled calmodulin (³⁵S-CaM); Coomassie-stained proteins were used as a loading control (Coomassie staining). **b**, Molecular characterization demonstrating the genotypes of the *Atsr1-1* plants complemented with wild type (cW) and three mutants (cM1, cM2, cM3) of *AtSR1* cDNA. See the legend of Supplementary Fig. 2a for definitions of PCR1 and PCR2. Anti-Flag: western blot detected with anti-Flag M2 monoclonal antibody. Twenty micrograms of total protein was loaded per lane. **c**, Phenotypic comparison; genotype is marked beneath the panel. **d**, Growth comparison in terms of fresh rosette weight. **e**, Disease resistance test. *Pst* DC3000 (OD₆₀₀ 0.0001) was infiltrated into the tested plants, and the c.f.u. were measured 3 days after infiltration. **f**, Quantification of free salicylic acid. Each result is the average of four replicates. **g**, Expression of *PR*, salicylic acid signalling and synthetic genes detected with northern blot. Genotypes are marked beneath, and probes to the right of the panel. All plants in these experiments were grown at 19–21 °C for 5 weeks; all data are presented as mean ± s.d. (*n* = 4).

ways[1–3], which set the plant defence system in a 'ready-to-go' state. This necessitates the involvement of negative regulators to achieve a tractable and balanced action in immune responses and prevent unnecessary overactivation, which usually results in retarded plant growth and even death[3,15,28]. The functional characterizations of AtSR1, and its interactions with the *EDS1* promoter and Ca²⁺/CaM, shed fresh light on how Ca²⁺ signalling conditions salicylic-acid-mediated plant immunity to reach a well-balanced defence against pathogens. Consistent with our findings, earlier reports have revealed that Ca²⁺ signalling also negatively affects plant defences, for example the action of Ca²⁺/CaM through MLO in barley[29]. Notably, Ca²⁺ signals generated through CNGC2 contribute to suppressing defence and maintaining normal plant growth[18,27]. How Ca²⁺-mediated signalling contributes to the establishment of plant immunity by modulating Ca²⁺/CaM-activated negative defence

regulators such as AtSR1 is a challenging question. Recent studies revealed that interference with or removal of negative regulators are effective approaches in activating plant defence[3,28,30]. Conceivably, interfering with the Ca²⁺ signal activating AtSR1, or AtSR1 itself, could also produce a quick activation of plant defence, a possibility that deserves further investigation.

METHODS SUMMARY

Arabidopsis knockout lines were obtained from *Arabidopsis* Biological Resource Center at Ohio State University. Knockout lines and primers for identification of homozygous knockout mutants are listed in Supplementary Table 1, and homozygous knockouts in single and double mutants were identified using polymerase chain reaction (PCR) analysis. *Pseudomonas syringae* was infiltrated into leaves of tested *Arabidopsis* plants for disease resistance assay or analysis of induced expression of *PR* genes. Hydrogen peroxide was detected with 3,3′-diaminobenzidine (DAB) staining, and autofluorescence was observed under a fluorescence microscope. Salicylic acid quantification was performed using high-performance liquid chromatography. Site-directed mutagenesis and deletion mutants of *AtSR1* were generated with PCR-based mutagenesis. Electrophoretic mobility-shift assay (EMSA) of DNA fragment, protoplast transient expression and chromatin immunoprecipitation were performed as described in Methods. Wild-type and mutated AtSR1 proteins were expressed in *Escherichia coli* for CaM overlay assay or expressed in wild-type and *Atsr1* background for functional analysis using the floral dip approach. Gene expression was detected with northern blot, RT–PCR or western blot analyses depending on specific purposes.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Lecourieux, D., Ranjeva, R. & Pugin, A. Calcium in plant defence-signalling pathways. *New Phytol.* **171**, 249–269 (2006).
2. Durrant, W. E. & Dong, X. Systemic acquired resistance. *Annu. Rev. Phytopathol.* **42**, 185–209 (2004).
3. Nimchuk, Z., Eulgem, T., Holt, B. F. III & Dangl, J. L. Recognition and response in the plant immune system. *Annu. Rev. Genet.* **37**, 579–609 (2003).
4. Yang, T. & Poovaiah, B. W. An early ethylene up-regulated gene encoding a calmodulin-binding protein involved in plant senescence and death. *J. Biol. Chem.* **275**, 38467–38473 (2000).
5. Reddy, A. S., Reddy, V. S. & Golovkin, M. A calmodulin binding protein from *Arabidopsis* is induced by ethylene and contains a DNA-binding motif. *Biochem. Biophys. Res. Commun.* **279**, 762–769 (2000).
6. Yang, T. & Poovaiah, B. W. A calmodulin-binding/CGCG box DNA-binding protein family involved in multiple signaling pathways in plants. *J. Biol. Chem.* **277**, 45049–45058 (2002).
7. Bouche, N., Scharlat, A., Snedden, W., Bouchez, D. & Fromm, H. A novel family of calmodulin-binding transcription activators in multicellular organisms. *J. Biol. Chem.* **277**, 21851–21861 (2002).
8. Han, J. *et al.* The fly CAMTA transcription factor potentiates deactivation of rhodopsin, a G protein-coupled light receptor. *Cell* **127**, 847–858 (2006).
9. Song, K. *et al.* The transcriptional coactivator CAMTA2 stimulates cardiac growth by opposing class II histone deacetylases. *Cell* **125**, 453–466 (2006).
10. Ryals, J. A. *et al.* Systemic acquired resistance. *Plant Cell* **8**, 1809–1819 (1996).
11. Lorrain, S., Vailleau, F., Balague, C. & Roby, D. Lesion mimic mutants: keys for deciphering cell death and defense pathways in plants? *Trends Plant Sci.* **8**, 263–271 (2003).
12. Alvarez, M. E. *et al.* Reactive oxygen intermediates mediate a systemic signal network in the establishment of plant immunity. *Cell* **92**, 773–784 (1998).
13. Bieri, S. *et al.* RAR1 positively controls steady state levels of barley MLA resistance proteins and enables sufficient MLA6 accumulation for effective resistance. *Plant Cell* **16**, 3480–3495 (2004).
14. Holt, B. F. III, Belkhadir, Y. & Dangl, J. L. Antagonistic control of disease resistance protein stability in the plant immune system. *Science* **309**, 929–932 (2005).
15. Heil, M. & Baldwin, I. T. Fitness costs of induced resistance: emerging experimental support for a slippery concept. *Trends Plant Sci.* **7**, 61–67 (2002).
16. Mauch, F. *et al.* Manipulation of salicylate content in *Arabidopsis thaliana* by the expression of an engineered bacterial salicylate synthase. *Plant J.* **25**, 67–77 (2001).
17. Yang, S. & Hua, J. A haplotype-specific resistance gene regulated by *BONZAI1* mediates temperature-dependent growth control in *Arabidopsis*. *Plant Cell* **16**, 1060–1071 (2004).
18. Clough, S. J. *et al.* The *Arabidopsis dnd1* 'defense, no death' gene encodes a mutated cyclic nucleotide-gated ion channel. *Proc. Natl Acad. Sci. USA* **97**, 9323–9328 (2000).
19. Petersen, M. *et al. Arabidopsis* map kinase 4 negatively regulates systemic acquired resistance. *Cell* **103**, 1111–1120 (2000).
20. Wiermer, M., Feys, B. J. & Parker, J. E. Plant immunity: the EDS1 regulatory node. *Curr. Opin. Plant Biol.* **8**, 383–389 (2005).

21. Nawrath, C., Heck, S., Parinthawong, N. & Metraux, J. P. EDS5, an essential component of salicylic acid-dependent signaling for disease resistance in *Arabidopsis*, is a member of the MATE transporter family. *Plant Cell* **14**, 275–286 (2002).

22. Wildermuth, M. C., Dewdney, J., Wu, G. & Ausubel, F. M. Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature* **414**, 562–565 (2001).

23. Xing, D. & Chen, Z. Effects of mutations and constitutive overexpression of EDS1 and PAD4 on plant resistance to different types of microbial pathogens. *Plant Sci.* **171**, 251–262 (2006).

24. Du, L. & Poovaiah, B. W. Ca^{2+}/calmodulin is critical for brassinosteroid biosynthesis and plant growth. *Nature* **437**, 741–745 (2005).

25. Gleason, C. *et al.* Nodulation independent of rhizobia induced by a calcium-activated kinase lacking autoinhibition. *Nature* **441**, 1149–1152 (2006).

26. Guo, F.-Q., Okamoto, M. & Crawford, N. M. Identification of a plant nitric oxide synthase gene involved in hormonal signaling. *Science* **302**, 100–103 (2003).

27. Ali, R. *et al.* Death don't have no mercy and neither does calcium: *Arabidopsis* cyclic nucleotide gated channel2 and innate immunity. *Plant Cell* **19**, 1081–1095 (2007).

28. Mackey, D., Belkhadir, Y., Alonso, J. M., Ecker, J. R. & Dangl, J. L. *Arabidopsis RIN4* is a target of the type III virulence effector *AvrRpt2* and modulates *RPS2*-mediated resistance. *Cell* **112**, 379–389 (2003).

29. Kim, M. C. *et al.* Calmodulin interacts with MLO protein to regulate defence against mildew in barley. *Nature* **416**, 447–451 (2002).

30. Shen, Q. H. *et al.* Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses. *Science* **315**, 1098–1103 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

METHODS

Identification of *Arabidopsis* knockout mutants and generation of double mutants. *Arabidopsis* homozygous knockout plants were isolated from plants germinated from the Salk T-DNA insertion collection using PCR analysis[31]. The effects of the T-DNA insert on the interrupted genes were confirmed by northern blot analysis or RT–PCR. The primers used for analysis of mutants are listed in Supplementary Table 1.

Arabidopsis AtSR1 homozygous knockout line *Atsr1-1* (Salk_001152) was crossed with *ics1/sid2* (Salk_088254), *pad4* (Salk_089936) and *eds5* (Salk_091541). Double mutants were selected from the F_2 population by PCR analysis.

***Pseudomonas syringae* infection, time course induction and disease resistance assay.** *Pseudomonas syringae* pv. *tomato* DC3000 culture and inoculation was performed as previously described[32,33]. Briefly, leaves of 4- to 5-week-old plants grown at $25–27\,^{\circ}C$ with 12-h photoperiod were infiltrated with *Pst* DC3000 at $OD_{600} = 0.001$ in 10 mM $MgCl_2$ for time course induction and disease resistance test. Leaves of 5-week-old plants grown at $19–21\,^{\circ}C$ with 12-h photoperiod were infiltrated with *Pst* DC3000 at $OD_{600} = 0.0001$ in 10 mM $MgCl_2$ for disease resistance assay. Each disease resistant result is the average of four replicates; the results are presented as mean ± s.d.

Detection of H_2O_2 and autofluorescence. *In situ* H_2O_2 detection was performed essentially as described earlier[34]. Leaves from wild-type and mutant plants were vacuum-infiltrated with 1 mg ml^{-1} DAB (Sigma). The infiltrated leaves were incubated in the DAB solution for 6 h under high-humidity conditions. Leaves were fixed and cleared of chlorophyll with several changes of a 1:3:1 mixture of lactic acid:ethanol:glycerol and mounted in Tris/glycerol and examined under a dissecting microscope for reddish-brown precipitate. Autofluorescence compounds were detected with a fluorescence microscope with a 488-nm excitation and 510-nm emission filter[35].

Measurement of salicylic acid through HPLC. Salicylic acid quantification was performed as previously described[36] with minor modification. Briefly, leaf tissue was collected from 5-week-old plants. For each sample, 150−200 mg tissue was ground in liquid nitrogen, and extracted with 90% methanol. After the extraction was dried, 500 μl 5% trichloroacetic acid was added to the residue. The free salicylic acid was extracted from the aqueous phase with ethylacetate-cyclopentane (1:1), and the organic phase was dried under nitrogen. The conjugated salicylic acid in the aqueous phase was hydrolysed at $100\,^{\circ}C$ in HCl solution with pH 1 for 30 min. The released free salicylic acid was extracted with organic mixture and dried as described above.

The dried extract was dissolved in 100 μl HPLC mobile phase, and 10 μl was injected into the HPLC column (SPHERISORD ODS-2, 4.6×150 mm, 5 μm, Waters), and chromatographic separation was performed at $40\,^{\circ}C$ with a flow rate of 1.0 ml min^{-1}. Salicylic acid was detected by a fluorescence detector.

Wild-type and mutated versions of *AtSR1* cDNA and −1.5-kb *EDS1* promoter. Full-length *AtSR1* cDNA was isolated from an *Arabidopsis* ZAP Express (Stratagene) by library screening. The full-length *AtSR1* cDNA was cut from the pBK-CMV vector with BamHI and XbaI and cloned into a modified pBluescript II KS+ vector without a SacI site for DNA manipulation. The 3′ region starting from the SacI site in *AtSR1* cDNA to the end of the coding sequence covering the CaM-binding domain was re-synthesized with PCR to generate site-directed and deletion mutants[32] (Supplementary Fig. 1b) and also to remove the stop codon and add an XbaI site for the purpose of fusing to the Flag tag; these fragments were used to replace the original 3′ region of *AtSR* cDNA clone to produce wild-type and mutated versions of *AtSR1* cDNA.

The −1.5-kb *EDS1* promoter region (*EDS1P*) was amplified from genomic DNA using EDS1P-F (5′-GCAAGCTTAGAGCTTTTAAGAATATTATGCACA-AGAGAGAG-3′) and EDS1P-R (5′-GCGGATCCTGATCTATATCTATTCTC-TTTTCTTTAGTGGACTTTCTT-3′) primers. Site-directed mutagenesis was used to change its ACGCGT (−746 to −741) to ACCCGT (*eds1p*)[32].

Preparation of expression constructs. The 3′ region (corresponding to amino acids 899–1032) of *AtSR1* covering the CaMBD (WSVGVLEKVILRWR-RKGAGLRGF) or the 5′ region of AtSR1 (corresponding to amino acids 1–153), or AtSR6 (corresponding to amino acids 1–158) covering the CG-1 DNA-binding domain were re-synthesized with PCR, and EcoRI and XhoI sites were added to the 5′ and 3′ ends of these fragments, and cloned into pET32A between its EcoRI and XhoI sites. The coding region of *ABF1* was amplified by PCR, BamHI and XhoI sites were added to the 5′ and 3′ ends of the fragments, and then they were cloned into pET32a.

The plant expression vector pDL28F is a derivative of pCambia1300 (AF234296) containing an extra cassette of '35S promoter-MCS-Flag-35S poly(A)' modified from pFF19 (ref. 37). Wild-type and mutated versions of

AtSR1 full-length cDNA without stop codons were cloned into pDL28F between its BamHI and XbaI sites and fused to its Flag tag. *NahG* (M60055) was also cloned into pDL28F but not fitted into the reading frame of Flag tag. The −1.5-kb *EDS1* promoter and its mutated version were digested with HindIII and BamHI and used to replace the 35S promoter in pDL28F. The coding region of luciferase (ABL09838) was amplified by PCR using Luc-F (5′-GCGGAT-CCATGGAAGACGCCAAAAACATAAAGAAAGG-3′) and Luc-R (5′-GCGT-CGACTTACAATTTGGACTTTCCGCCCTTCTTGG-3′) primers, and cloned into pDL28F/EDS1P to produce *EDS1P::luc* (pDL326) or into pDL28F/eds1p to produce *eds1p::luc* (pDL327) constructs. The flower dipping approach was used to deliver plant expression constructs for stable transformation.

Recombinant protein purification, EMSA and ^{35}S-CaM-binding assay. The *E. coli* strain BL21(DE3)/pLysS carrying the above pET32a-derived plasmids for expression of recombinant proteins of the wild-type and mutated versions of AtSR1, AtSR6 or ABF1 were induced with 0.5 mM IPTG for 3 h. 6_{His}-tagged recombinant proteins were purified using Ni-NTA agarose affinity beads (Qiagen) as described by the manufacturer. Recombinant AtSR1 or AtSR6 covering CG-1 domain, or ABF1, was used for EMSA[32] to detect its interaction with wild-type or mutated *EDS1* promoter fragments. Recombinant AtSR1 containing CaMBD was used for CaM overlay assay[38].

Chromatin immunoprecipitation analysis. Thirty million leaf mesophyll protoplasts from 4-week-old wild-type *Arabidopsis* plants were transfected with 60 μg of YFP control DNA or *AtSR1-YFP* DNA with the PEG-mediated transformation method[39]. Protoplasts were incubated at $25\,^{\circ}C$ under dark conditions for 16 h before chromatin immunoprecipitation assay[40]. The harvested cells were resuspended in W5 medium containing 1% formaldehyde and crosslinked for 20 min. The protoplasts were lysed and DNA was sheared on ice with sonication. The pre-cleared lysate was incubated with 60 μl anti-GFP agarose beads (D153-8, MBL) for 12 h at $4\,^{\circ}C$. Beads were washed five times, re-suspended in elution buffer and incubated at $65\,^{\circ}C$ for 12 h. After purification, the DNA was amplified with PCR using EDSUE-F (5′-TGGCTTTTCGTAGAAATTTCCC-3′) and EDSUE-R (5′-GGAACCGGTTCGATTTCTCTC-3′) primers.

***eds1::luc* transient expression assays.** One-million protoplasts from 4- to 5-week-old wild-type and *Atsr1-1* plants grown at $20\,^{\circ}C$ were transfected in four replicates with 5 μg GUS plasmid (as internal control) and 5 μg pDL326 or pDL327 plasmids with the PEG-mediated transfection method[39]. After 16–17 h incubation at $20\,^{\circ}C$, the protoplasts were harvested and luciferase assays were performed using a luciferase assay kit (Promega). To account for variation in transfection efficiencies, GUS assays were performed with each treatment using standard methyl umbelliferyl glucoronide substrate. The data presented are the average of the Luc/GUS ratios of four replications ± s.d.

31. Alonso, J. M. *et al.* Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* **301**, 653–657 (2003).
32. Du, L. & Chen, Z. Identification of genes encoding receptor-like protein kinases as possible targets of pathogen- and salicylic acid-induced WRKY DNA-binding proteins in *Arabidopsis*. *Plant J.* **24**, 837–847 (2000).
33. Chen, K., Du, L. & Chen, Z. Sensitization of defense responses and activation of programmed cell death by a pathogen-induced receptor-like protein kinase in *Arabidopsis*. *Plant Mol. Biol.* **53**, 61–74 (2003).
34. Thordal-Christensen, H., Zhang, Z., Wei, Y. & Collinge, D. B. Subcellular localization of H_2O_2 in plants. H_2O_2 accumulation in papillae and hypersensitive response during the barley—powdery mildew interaction. *Plant J.* **11**, 1187–1194 (1997).
35. Yu, I. C., Parker, J. & Bent, A. F. Gene-for-gene disease resistance without the hypersensitive response in *Arabidopsis dnd1* mutant. *Proc. Natl Acad. Sci. USA* **95**, 7819–7824 (1998).
36. Clarke, J. D., Volko, S. M., Ledford, H., Ausubel, F. M. & Dong, X. Roles of salicylic acid, jasmonic acid, and ethylene in cpr-induced resistance in *Arabidopsis*. *Plant Cell* **12**, 2175–2190 (2000).
37. Timmermans, M. C., Maliga, P., Vieira, J. & Messing, J. The pFF plasmids: cassettes utilising CaMV sequences for expression of foreign genes in plants. *J. Biotechnol.* **14**, 333–344 (1990).
38. Du, L. & Poovaiah, B. W. A novel family of Ca^{2+}/calmodulin-binding proteins involved in transcriptional regulation: Interaction with fsh/Ring3 class transcription activators. *Plant Mol. Biol.* **54**, 549–569 (2004).
39. Yoo, S.-D., Cho, Y.-H. & Sheen, J. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols* **2**, 1565–1572 (2007).
40. Hernandez, J. M., Feller, A., Morohashi, K., Frame, K. & Grotewold, E. The basic helix loop helix domain of maize R links transcriptional regulation and histone modifications by recruitment of an EMSY-related factor. *Proc. Natl Acad. Sci. USA* **104**, 17222–17227 (2007).

Direct observation of the nanoscale dynamics of membrane lipids in a living cell

Christian Eggeling[1]*, Christian Ringemann[1]*, Rebecca Medda[1], Günter Schwarzmann[2], Konrad Sandhoff[2], Svetlana Polyakova[1], Vladimir N. Belov[1], Birka Hein[1], Claas von Middendorff[1], Andreas Schönle[1] & Stefan W. Hell[1]

Cholesterol-mediated lipid interactions are thought to have a functional role in many membrane-associated processes such as signalling events[1–5]. Although several experiments indicate their existence, lipid nanodomains ('rafts') remain controversial owing to the lack of suitable detection techniques in living cells[4,6–9]. The controversy is reflected in their putative size of 5–200 nm, spanning the range between the extent of a protein complex and the resolution limit of optical microscopy. Here we demonstrate the ability of stimulated emission depletion (STED) far-field fluorescence nanoscopy[10] to detect single diffusing (lipid) molecules in nanosized areas in the plasma membrane of living cells. Tuning of the probed area to spot sizes ∼70-fold below the diffraction barrier reveals that unlike phosphoglycerolipids, sphingolipids and glycosylphosphatidylinositol-anchored proteins are transiently (∼10–20 ms) trapped in cholesterol-mediated molecular complexes dwelling within <20-nm diameter areas. The non-invasive optical recording of molecular time traces and fluctuation data in tunable nanoscale domains is a powerful new approach to study the dynamics of biomolecules in living cells.

Unlike phosphoglycerolipids, sphingolipids and glycosylphosphatidylinositol (GPI)-anchored proteins are assumed to form molecular complexes or integrate, assisted by cholesterol, into <200-nm-sized lipid nanodomains impeding their diffusion in the plasma membrane[11]. Featuring a diameter $d > 200$ nm, the commonly used detection spot of a confocal microscope averages over such details of molecular diffusion (Fig. 1a). Electron microscopy provides the required resolution, but not in living cells[12]. Atomic force microscopy[13], near-field optical microscopy[14] and Förster resonance energy transfer[15] could in principle provide an answer, but the direct visualization of lipid nanodomains rather than that of clustered proteins is challenged by the rapid diffusion of the lipids[4]. This rapid diffusion also prevents the fast tracking of single lipid molecules in their native cargo-free state[16–18]. Although fluorescence recovery after photobleaching[19,20] and confocal fluorescence correlation spectroscopy[21–23] are able to map the diffusion, their resolution is diffraction-limited which requires, for example, mechanical contact with nanometric apertures[24]. In contrast, emergent stimulated emission depletion (STED) fluorescence microscopy[10,25,26] (see Methods) opens up a new avenue to determine lipid diffusion, because its detection area can uniquely be downscaled in size by suppressing the fluorophore excitation at the outer parts of the focal spot (Fig. 1b).

Therefore we labelled sphingolipids, represented by sphingomyelin and ganglioside GM1, a GPI-anchor as well as the putatively 'non-integrating' phosphoethanolamine with the fluorophore Atto647N and studied their dynamics in the plasma membrane of living cells (see Methods). Control experiments indicated that the dye label had

no observable influence on the dynamic behaviour of the lipids or the GPI-anchor (Supplementary Fig. 8). An average of <10 labelled molecules per μm² membrane area ensured the presence of single fluorescent molecules in the detection area. The measurements were performed in living mammalian cells under physiological conditions. Specifically, we targeted arbitrary points on the plasma membrane and recorded the fluorescence bursts of molecules crossing the focal spot. Figure 1c and e show single-molecule fluorescence time traces of phosphoethanolamine, measured using the regular confocal and the subdiffraction detection area downsized by STED to $d \approx 40$ nm, respectively. Besides demonstrating the ability of the STED microscope to detect single molecules in living cells, the traces show a uniform reduction of the burst length, accounting for the reduced transit time of a phosphoethanolamine molecule through the subdiffraction sized spot.

Next we recorded a similar pair of confocal and STED traces of sphingomyelin molecules (Fig. 1d and f). The confocal measurement does not show obvious differences between the sphingomyelin and phosphoethanolamine diffusion. Containing ∼1,000 photons per molecule at an average transit time of 10 ms, the fluorescence bursts are rather homogeneous in height and length for both lipids. In contrast, the STED microscope reveals a whole range of burst durations for sphingomyelin and hence a substantial difference to phosphoethanolamine. Whereas the phosphoethanolamine traces feature sharp peaks only, in the case of sphingomyelin the sharp peaks are accompanied by longer bursts. Thus the reduction of the detection area far below that provided by confocal microscopy reveals that the diffusion of sphingomyelin is strongly heterogeneous.

An important observation is that STED microscopy reduces the fluorescence burst length but hardly its peak height. The detected photon count-rate of a typical molecular trace peaks at 100 kHz, which enables the analysis of STED traces with a good signal-to-noise ratio. Hence we plotted a frequency histogram displaying the number of counts along with the burst duration for about 500 bursts recorded for phosphoethanolamine and sphingomyelin (Fig. 1g, h and Supplementary Fig. 2). For phosphoethanolamine we found just a single type of diffusion, characterized by burst durations <1 ms. Phosphoethanolamine shows no observable sign of heterogeneous diffusion in the plasma membrane. In contrast, the sphingomyelin histogram reveals a broad distribution of burst durations from short events that are similar to that of phosphoethanolamine, to longer events with burst durations ranging up to >50 ms.

To quantify the lipid dynamics further, we applied fluorescence correlation spectroscopy (FCS) averaging over many traces from individual molecules diffusing in and out of a well-known focal spot[27]. The decay of the correlation curve yields the average molecular transit time

[1]Department of Nanobiophotonics, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany. [2]LIMES Membrane Biology and Lipid Biochemistry Unit, University of Bonn, Gerhard-Domagk-Strasse 1, 53121 Bonn, Germany.
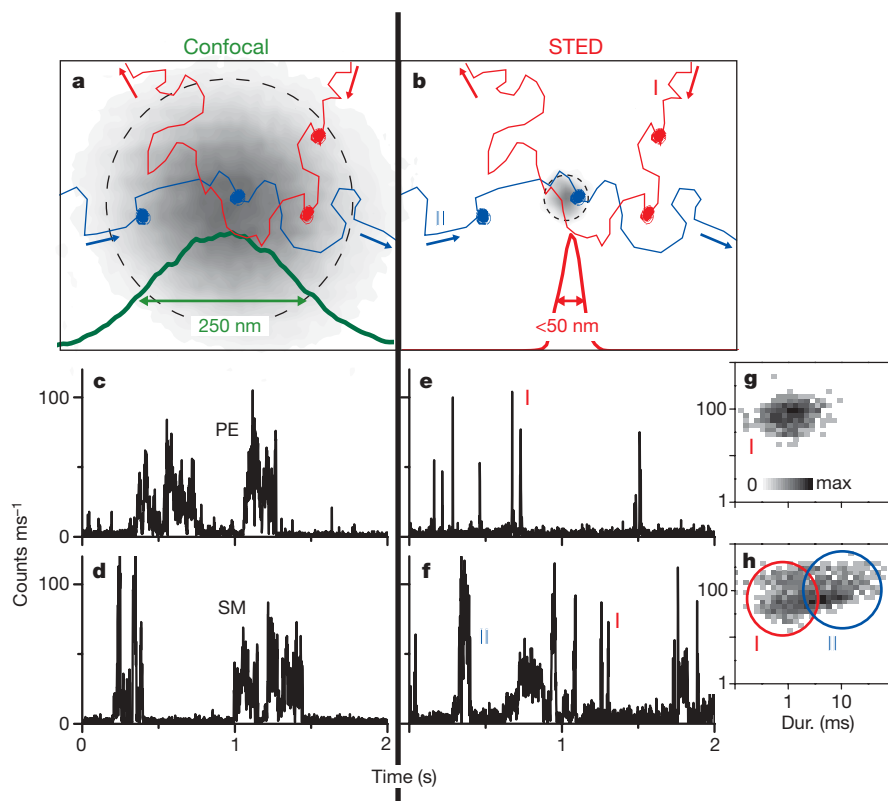*These authors contributed equally to this work.

Figure 1 | STED microscopy time traces of single-molecule diffusion in live cell plasma membrane. **a**, **b**, Molecules may move freely and/or be transiently trapped on small spatial scales. The large detection area of a confocal microscope (**a**) cannot discern such details. However, the subdiffraction spot created by STED (**b**) is able to discriminate between lipids that diffuse freely (I) and those that are hindered (II) during their passage. **c**–**h**, Fluorescence bursts from single-diffusing Atto647N-labelled phosphoethanolamine (PE) and sphingomyelin (SM) lipids detected with a confocal (**c**, **d**) and a STED spot (**e**, **f**), and a frequency plot of the value pairs of fluorescence counts per millisecond and duration (dur.) of selected bursts of the STED recording for phosphoethanolamine (**g**, 497 bursts) and sphingomyelin (**h**, 539 bursts). By ensuring that a freely diffusing molecule spends only little time in the detection area, STED microscopy distinguishes hindered (II) from free diffusion (I).

through the focal area (Supplementary Information). The confocal correlation curves recorded for sphingomyelin are only slightly shifted compared to those recorded for phosphoethanolamine (Fig. 2a). Because both curves can be described by normal diffusion of a single species, we cannot infer whether the diffusion of sphingomyelin is inhomogeneous on small spatial scales, or just slightly slower than phosphoethanolamine. This is different in the correlation curves recorded for $d \approx 40$ nm using STED, because here, marked disparities between phosphoethanolamine and sphingomyelin diffusion become apparent (Fig. 2b). Owing to the smaller detection area, the phosphoethanolamine curve is shifted towards shorter time scales, but still represents a single normal diffusion. In contrast, the sphingomyelin curve now features much longer time scales, and can be fitted only by a model assuming two dissimilar modalities of focal transits or, alternatively, anomalous diffusion of the lipids (Supplementary Information). To investigate whether the longer focal transit time of sphingomyelin is associated with cholesterol, we added cholesterol oxidase to deplete the cholesterol in the membrane. Figure 2c shows that the longer transit times disappear.

STED microscopy enabled us to scale the diameter of the focal detection area from $d = 250$ nm down to 30 nm after increasing the intensity of the STED light (Supplementary Fig. 1). The accompanying decrease in the transit time reveals the multifaceted dynamics of the membrane molecules[23,24]. The transit time for phosphoethanolamine decreases in proportion to the ~70-fold reduction in focal detection area (Fig. 3), indicating that phosphoethanolamine is indeed freely diffusing with a diffusion coefficient $D = (5 \pm 2)\ 10^{-9}\ \mathrm{cm^2\,s^{-1}}$. This is different for sphingomyelin. For a relatively large subdiffraction spot diameter $d > 160$ nm, the diffusion of sphingomyelin appears normal and just 1.5–2.5 times slower than that of phosphoethanolamine.

Likewise, for $80\ \mathrm{nm} < d < 160\ \mathrm{nm}$, the spot size is not small enough to safely quantify anomalous diffusion of sphingomyelin. However, for $d < 80$ nm, we can separate free from hindered diffusion and analyse the sphingomyelin correlation data by applying a model of two distinct modalities of focal transits: free and hindered diffusion. Considering that the free diffusion exhibits the same transit time τ_{D1} as phosphoethanolamine (compare Fig. 1g and h), we calculated that a fraction of $A_2 \approx 60\%$ of sphingomyelin molecules crosses the $d = 30$–80 nm large areas within a transit time τ_{D2} which is >10 times longer than τ_{D1}, reaching a plateau at ~10 ms.

This observation of a restricted minimum transit time through nanoscale areas rules out the possibility that sphingomyelin diffuses freely (but just more slowly) in the plasma membrane, because the hallmark of free diffusion is a transit time that scales with d^2; compare τ_{D1}. Therefore the hindered diffusion of sphingomyelin and the longer transit time τ_{D2} must be due to a brief trapping of these molecules. The average trapping period $\tau_{\mathrm{trap}} \sim 10$ ms is calculated by subtracting τ_{D1} from τ_{D2}. The fact that τ_{trap} is constant for $d < 60$ nm indicates a single trapping event within the probed areas. Similar trapping was observed for the Atto647N-labelled GPI-anchor (Fig. 3) and the ganglioside GM1 (Supplementary Fig. 6) with $D \approx 3$ and $5 \times 10^{-9}\ \mathrm{cm^2\,s^{-1}}$, $\tau_{\mathrm{trap}} \approx 18$ and 11 ms, and $A_2 \approx 35$ and 45%, respectively.

Notably, the addition of the cholesterol-depleting agent cholesterol oxidase causes sphingomyelin, the GPI-anchor, and GM1 to exhibit a linear dependence of the transit time on d^2, which is indicative of the fact that the trapping has been reduced. Although cholesterol oxidase may also induce other changes in the membrane, it primarily depletes the cholesterol. Alternative cholesterol depletion by β-cyclodextrin also reduces the trapping (Supplementary Fig. 5).
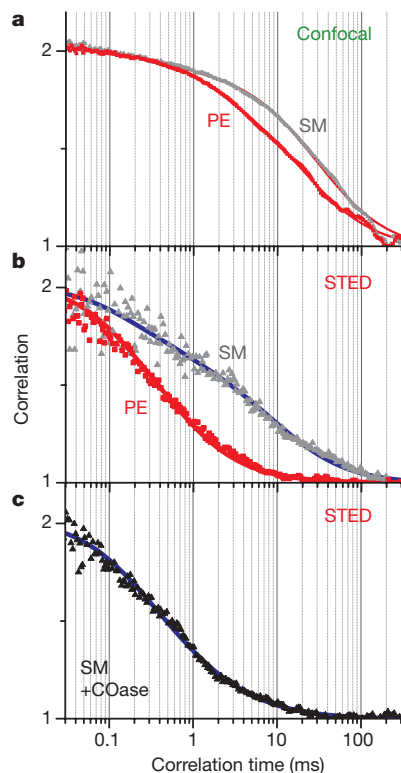
Figure 2 | FCS of Atto647N-labelled phosphoethanolamine and sphingomyelin plasma membrane diffusion. a–c, Normalized correlation data of phosphoethanolamine (PE, red dots) and sphingomyelin (SM, grey dots) for confocal (**a**) and STED (**b**) recording, and of sphingomyelin (**c**) after cholesterol depletion by the addition of cholesterol oxidase (COase, black dots) for STED recording. The STED but not the confocal data reveal cholesterol-assisted heterogeneous diffusion of sphingomyelin. Red lines in **a** and **b** denote single-species fit; $\tau_D = 19$ (phosphoethanolamine, confocal), 28 (sphingomyelin, confocal) and 0.45 ms (phosphoethanolamine, STED). Blue lines denote fit assuming two dissimilar modalities of focal transits: $\tau_{D1} = 0.45$ ms, $\tau_{D2} = 10$ ms and $A_2 = 64\%$ (sphingomyelin, in **b**) and $\tau_{D2} = 4$ ms and $A_2 = 15\%$ (sphingomyelin + cholesterol oxidase, in **c**).

The hindered sphingolipid and GPI-anchor diffusion may be caused by various molecular interactions. However, we can exclude the floating of temporally stable lipid complexes or domains through the membrane, because this would manifest itself as a linear dependence of τ_{D2} on the reduced focal area d^2 created by STED. Moreover, our data demonstrate that during the trapping, the sphingolipids or GPI-anchored molecules remain within a <20-nm diameter area. This conclusion is on the basis of the fact that for 30–60 nm large spots, the transit time τ_{D2} and the fraction A_2 of diffusion-hindered events is constant (Supplementary Information). These findings are supported by alternative evaluations of the FCS data applying anomalous diffusion models (Supplementary Fig. 3), by the STED images recorded for sphingomyelin (Supplementary Fig. 10), and by Monte Carlo simulations of hindered lipid dynamics (Supplementary Fig. 9).

Altogether, the evolving picture of the observed sphingolipid and GPI-anchor dynamics is that of transient formations of cholesterol-assisted molecular complexes, such as lipid-protein binding or lipid shells (compare Supplementary Information). Although an entire complex of for example several lipid and protein molecules may (temporally) be of larger spatial scales, the trapped molecule dwells only within <20-nm diameter areas. The diffusion of the complexes must be slow, because they must diffuse through the focal spot within a time period that is longer than the trapping time, that is, with a diffusion constant $D \ll 0.2 \times d^2/\tau_{trap} \approx 10^{-10}\,\mathrm{cm^2\,s^{-1}}$.

The linear dependence of the transit time of the freely diffusing lipids on the subdiffraction detection area (and their average particle number per area, Supplementary Fig. 11) proves that our data are not
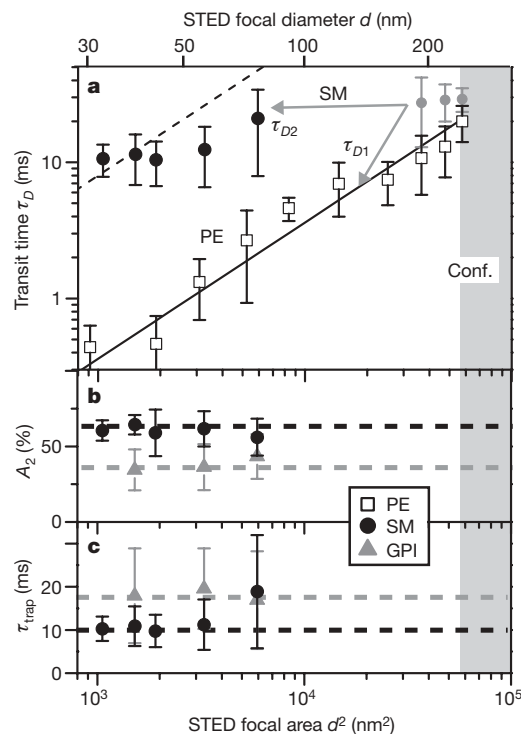

Figure 3 | Molecular transit through nanoscale areas in live cell plasma membrane. The parameters measured by STED-FCS in focal detection areas ranging from $d = 250$ nm (conf., confocal; grey area) down to 30 nm in diameter. **a,** The transit times τ_D determined for phosphoethanolamine (PE, open squares) decrease linearly with the area d^2, confirming free diffusion (solid line). Whereas large areas >180 nm in diameter yield just a single transit time τ_D of sphingomyelin (SM, grey dots), detection areas <80 nm in diameter reveal two distinct modalities of molecular transits (grey arrows), with $\tau_{D1} = \tau_D$(phosphoethanolamine) (solid line) and τ_{D2} demonstrating hindered diffusion. The τ_{D2} plateau found for small d indicates transient trapping of sphingomyelin. Slowed-down free diffusion of sphingomyelin can be ruled out because such behaviour would display linear dependence of τ_D on d^2, as exemplified for $D = 2 \times 10^{-10}\,\mathrm{cm^2\,s^{-1}}$ (dashed line). **b, c,** Fraction A_2 (**a**) and trapping time τ_{trap} (**c**) of transits of sphingomyelin (black dots) and GPI-anchor (grey triangles) exhibiting transient trapping. Standard deviations given as error bars result from averaging over more than thirty FCS measurements on different cells.

compromised by photobleaching, local heating or radical formation. Even if some molecules were bleached on passing the doughnut of STED light, the diffusion of those that have entered the detection area is less affected. In fact, the confocal measurements are more prone to bleaching artefacts due to the inherently longer transit time through the detection area. Thus the traces in Figs 1 and 2 show the ability of STED microscopy to provide a unique and hitherto unrecognized access to the dynamics of single molecules in nanosized areas in living cells, thereby greatly expanding the potential of the popular FCS technique. STED-FCS is complementary to single-molecule tracking because several molecules per subdiffraction area may contribute to the signal, allowing a fast gathering of statistically accurate data. Besides, the spot size of a STED microscope is physically predefined and the spot can be directed to arbitrary coordinates within a cell. All these features have provided a unique combination of temporal and spatial resolution required to quantify the nanoscale dynamics of the small lipid molecules in the cellular membrane. Thus the trapped molecular diffusion revealed in this work is only a prelude of a new class of nanoscale biomolecular studies that are to follow.

METHODS SUMMARY

In our STED microscope, the diffraction-limited focal spot of excitation light is overlapped with a doughnut-shaped spot of STED light that switches off the ability of the label to fluoresce, thus confining the origin of the fluorescence to

1161

the doughnut centre (Supplementary Fig. 1). With the intensity I_s being a characteristic of the fluorophore and the STED light, and $I \gg I_s$ denoting the doughnut peak intensity, the diameter of the detection area can be tuned as $d \approx \lambda / \left(2\mathrm{NA} \sqrt{1 + I/I_s} \right) \ll 250\,\mathrm{nm}$, with NA denoting the numerical aperture of the lens, and λ the wavelength used for STED[25,26]. We used a 76-MHz train of 250–300 ps pulses for STED at 770–780 nm, and an average focal power of up to 380 mW. Imprinting a helical phase delay on the STED laser wavefront rendered a doughnut-shaped focal spot. The STED pulses triggered a 633-nm pulsed laser diode for fluorescence excitation with average powers of 5–8 µW. Laser focusing and collection of the fluorescence was performed by an NA 1.42 oil immersion lens. The signal was detected by avalanche photo diodes together with a correlator or single-photon counting hardware. The experiments were run and analysed using Imspector as well as self-written software. The focal diameters d of the STED microscope were calibrated by imaging 20-nm diameter fluorescent beads (Supplementary Fig. 1). Labelling of the lipids with the organic dye Atto647N was accomplished by chemical synthesis and of the GPI-anchor by acyl carrier protein (ACP)-tagging, respectively. Incubation with the fluorescent lipid–BSA complexes or GPI–ACP plasmid transfection resulted in molecular insertion into the plasma membrane of living PtK2 cells. Measurements were performed in cell culture medium at 27–37 °C, using a temperature-controlled sample chamber and objective lens heater, and by placing the foci of the inverted microscope randomly on the lower plasma membrane.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Simons, K. & Ikonen, E. Functional rafts in cell membranes. *Nature* **387**, 569–572 (1997).
2. Brown, D. A. & London, E. Structure and function of sphingolipid- and cholesterol-rich membrane rafts. *J. Biol. Chem.* **275**, 17221–17224 (2000).
3. Fielding, C. J. *Lipid Rafts and Caveolae* (Wiley-VCH, 2006).
4. Jacobson, K., Mouritsen, O. G. & Anderson, G. W. Lipid rafts: at a crossroad between cell biology and physics. *Nature Cell Biol.* **9**, 7–14 (2007).
5. Hanzal-Bayer, M. F. & Hancock, J. F. Lipid rafts and membrane traffic. *FEBS Lett.* **581**, 2098–2104 (2007).
6. Munro, S. Lipid rafts: elusive or illusive? *Cell* **115**, 377–388 (2003).
7. Lommerse, P. H. M., Spaink, H. P. & Schmidt, T. *In vivo* plasma membrane organization: results of biophysical approaches. *Biochim. Biophys. Acta* **1664**, 119–131 (2004).
8. Hancock, J. F. Lipid rafts: contentious only from simplistic standpoints. *Nature Rev. Mol. Cell Biol.* **7**, 456–462 (2006).
9. Shaw, A. S. Lipid rafts: now you see them, now you don't. *Nature Immunol.* **7**, 1139–1142 (2006).
10. Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated emission depletion microscopy. *Opt. Lett.* **19**, 780–782 (1994).
11. Pike, L. J. Rafts defined: a report on the Keystone symposium on lipid rafts and cell function. *J. Lipid Res.* **47**, 1597–1598 (2006).
12. Fujita, A. *et al.* Gangliosides GM1 and GM3 in the living cell membrane form clusters susceptible to cholesterol depletion and chilling. *Mol. Biol. Cell* **18**, 2112–2122 (2007).
13. Binnig, G., Quate, C. F. & Gerber, C. Atomic force microscope. *Phys. Rev. Lett.* **56**, 930–933 (1986).
14. Pohl, D. W., Denk, W. & Lanz, M. Optical stethoscopy: Image recording with resolution $\lambda/20$. *Appl. Phys. Lett.* **44**, 651–653 (1984).
15. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913–916 (2002).
16. Saxton, M. J. & Jacobson, K. Single particle tracking: applications to membrane dynamics. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 373–399 (1997).
17. Schütz, G. J., Kada, G., Pastushenko, V. P. & Schindler, H. Properties of lipid microdomains in a muscle cell membrane visualized by single molecule microscopy. *EMBO J.* **19**, 892–901 (2000).
18. Fujiwara, T., Ritchie, K., Murakoshi, H., Jacobson, K. & Kusumi, A. Phospholipids undergo hop diffusion in compartmentalized cell membrane. *J. Cell Biol.* **157**, 1071–1081 (2002).
19. Yechiel, E. & Edidin, M. Micrometer-scale domains in fibroblast plasma membranes. *J. Cell Biol.* **105**, 755–760 (1987).
20. Feder, T. J., Brust-Mascher, I., Slattery, J. P., Baird, B. A. & Webb, W. W. Constrained diffusion or immobile fraction on cell surfaces: a new interpretation. *Biophys. J.* **70**, 2767–2773 (1996).
21. Fahey, P. F. *et al.* Lateral diffusion in planar lipid bilayers. *Science* **195**, 305–306 (1977).
22. Schwille, P., Korlach, J. & Webb, W. W. Fluorescence correlation spectroscopy with single-molecule sensitivity on cell and model membranes. *Cytometry* **36**, 176–182 (1999).
23. Wawrezinieck, L., Rigneault, H., Marguet, D. & Lenne, P.-F. Fluorescence correlation spectroscopy: diffusion laws to probe the submicron cell membrane organization. *Biophys. J.* **89**, 4029–4042 (2005).
24. Wenger, J. *et al.* Diffusion analysis within single nanometric apertures reveals the ultrafine cell membrane organization. *Biophys. J.* **92**, 913–919 (2007).
25. Willig, K. I., Rizzoli, S. O., Westphal, V., Jahn, R. & Hell, S. W. STED-microscopy reveals that synaptotagmin remains clustered after synaptic vesicle exocytosis. *Nature* **440**, 935–939 (2006).
26. Hell, S. W. Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007).
27. Magde, D., Elson, E. L. & Webb, W. W. Thermodynamic fluctuations in a reacting system - measurement by fluorescence correlation spectroscopy. *Phys. Rev. Lett.* **29**, 705–708 (1972).

METHODS

Cell culture. The epithelial cell line PtK2 was grown as previously described[28]. The cells were seeded on number 1 thickness standard glass coverslips (diameter 25 mm) to a confluence of about 80% and grown at 37 °C in a water-saturated atmosphere of 5% CO_2 in air. All media and supplements were purchased from Invitrogen.

Atto647N-labelled lipids. We used the organic dye Atto647N (fluorescence excitation and emission maxima at 645 and 670 nm, respectively; Atto-Tec) as a fluorescence marker. N-(Atto647N)-1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine (Atto647N-phosphoethanolamine, referred to as phosphoethanolamine) and N-(Atto647N)-sphingosylphosphocholine (N-Atto647-sphingomyelin, Atto647N-sphingomyelin, referred to as sphingomyelin) were purchased from Atto-Tec. The Atto647N-labelled ganglioside GM1 was synthesized as outlined in the Supplementary Information starting from GM1 (ALEXIS Biochemicals). We performed a series of control experiments using lipids labelled at different positions either with lipophilic or hydrophilic fluorescent labels. The control experiments together with the structures of all lipids are depicted in Supplementary Figs 7 and 8. The experiments indicate no observable influence of the used label on the dynamic characteristics of the labelled lipids.

Plasma membrane insertion of the fluorescent lipids. Complexes of the labelled lipids and bovine serum albumin (BSA) were prepared and incubated with the cells according to a slightly modified protocol than that previously described[29,30]; as detailed in the Supplementary Information.

Glycosylphosphatidylinositol anchor. We used an ACP tag (Covalys Bioscience) for labelling of the GPI-anchor with the organic dye Atto647N. Twenty-four hours after transfection with the plasmid pAEMXT-ACPwt-GPI (Covalys), we incubated PtK2 cells expressing GPI-anchored ACP for 30 min at 37 °C in complete medium supplemented with 5 µM CoA-Atto647N (AttoTec; CoA, coenzyme A), 5 mM $MgCl_2$ (Fluka) and 1 µM ACP synthase (Covalys). Afterwards, the cells were washed three times and prepared for measurements using DMEM containing no phenol red. The cells were measured 24 h after transfection. Control experiments with different fluorescent markers precluded an observable influence of the label on GPI trapping (Supplementary Information).

Treatment with cholesterol oxidase and β-cyclodextrin. To modify the cholesterol contents of the plasma membrane, the cells were either treated with 1 U ml^{-1} *Streptomyces* sp. cholesterol oxidase (Sigma, stock solution 34 U ml^{-1} in 50 mM KH_2PO_4, pH 7.5) in HDMEM (HEPES + DMEM) and then washed in HDMEM (see Supplementary Information), or treated with 10 mM β-cyclodextrin (Sigma) in HDMEM (without phenol red) for 30 min at 37 °C. The cholesterol oxidase or β-cyclodextrin treatment was performed either before or after the insertion of the fluorescent lipid analogues into the plasma membrane.

STED microscope. We implemented a confocalized setup in which the excitation of the dye was performed with a 633 nm pulsed laser diode ($\approx$ 80 ps pulse width, LDH-P-635, Picoquant). The STED beam was provided by a titanium:sapphire laser system (MaiTai, Spectra-Physics) operating at 770–780 nm with a repetition rate of 76 MHz; this beam also provided the trigger for the excitation laser. The STED laser power was controlled and stabilized by a power controller unit (LPC, Brockton Electronics). After passing a 30-cm optical glass rod, the linear-polarized light was coupled into a 120 m long polarization maintaining single-mode fibre (AMS Technologies) for stretching the pulses to 250–300 ps. The pulse timing of both lasers was adjusted using a home-built electronic delay unit. After spatial overlay of both laser beams with appropriate dichroic filters (AHF Analysentechnik) they were directed to a beam-scanning device (mirror

tilting system PSH 10/2, Piezosystem Jena) and then directed into the microscope (DMIRBE, Leica Microsystems). We used an oil immersion objective (PLAPON ×60, NA = 1.42, Olympus) to focus the laser light to a spot on the sample and to collect the fluorescence. The beam-scanning device allowed an exact control of the lateral position of the focal spots on the sample and enabled scanning. The axial position of the focal spots was adjusted by an objective lens positioning system (MIPOS 250, Piezosystem Jena). The doughnut-shaped focal spot of the STED beam featuring a central zero intensity was produced by introducing a phase-modifying plate (RPC Photonics) into the beam path, imprinting on the wave front a helical phase ramp $\exp(i\varphi)$ with $0 \leq \varphi \leq 2\pi$. A $\lambda/4$-plate ensured circular polarization of the STED and of the excitation beam[31,32]. The fluorescence was imaged back over the beam-scanning device and coupled into a multi-mode fibre splitter (Fibre Optic Network Technology) with an aperture size corresponding to ×1.4 the magnified excitation spot. The 50:50 split fluorescence signal was then detected by two single-photon counting units (avalanche photo diode SPCM-AQR-13-FC, Perkin Elmer Optoelectronics). We blocked potential contributions from scattered laser light or unwanted autofluorescence by appropriate emission filters (AHF Analysentechnik). The fluorescence counts were further processed by a hardware correlator card (Flex02-01D, Correlator.com) for FCS measurements or by a single-photon counting PC card (SPC 830, Becker & Hickl GmbH) for direct single-molecule analysis. The focal intensity distribution of the excitation and STED light were measured by scanning a scattering gold bead of sub-diffraction diameter (80 nm gold colloid, En.GC80, BBinternational) using a non-confocal detector (MP 963 Photon Counting Module, Perkin Elmer). The (pulse) intensity of the excitation light was 3–5 MW cm^{-2} ($\sim$25 kW cm^{-2} on average), stemming from $P = 5$–8 µW on the confocal spot of diameter (full-width-at-half maximum) $d = 250$ nm.

Measurements. We analysed the lipid dynamics by placing the foci on random positions in the lower plasma membrane facing the coverslip and completed all measurements before disruptive internalization, or any morphological changes in the cell could take place. The measurement times were kept short ($\sim$15 s) to avoid biasing distortion of the correlation data due to very infrequent transits of bright particles such as cell debris. The coverslips were mounted in a special microscope chamber (RC-40, Warner Instruments). This chamber together with an objective heater (Bioptechs Inc.) allowed a precise control of the sample's temperature. We performed most measurements at 27–37 °C in HDMEM as outlined in detail in the Supplementary Information. We excluded diffusion of non-integrated lipids (or dye tags) by control measurements in between the cells.

28. Osborn, M., Franke, W. W. & Weber, K. Visualization of a system of filaments 7–10 nm thick in cultured cells of an epithelioid line (PtK2) by immunofluorescence microscopy. *Proc. Natl Acad. Sci. USA* **74**, 2490–2494 (1977).

29. Martin, O. C. & Pagano, R. C. Internalization and sorting of a fluorescent analogue of glucosylceramide to the Golgi apparatus of human skin fibroblasts: utilization of endocytic and nonendocytic transport mechanisms. *J. Cell Biol.* **125**, 769–781 (1994).

30. Schwarzmann, G., Hofmann, P., Pütz, U. & Albrecht, B. Demonstration of direct glycosylation of nondegradable glucosylceramide analogs in cultured cells. *J. Biol. Chem.* **270**, 21271–21276 (1995).

31. Keller, J., Schönle, A. & Hell, S. W. Efficient fluorescence inhibition patterns for RESOLFT microscopy. *Opt. Express* **15**, 3361–3371 (2007).

32. Willig, K. I. *et al.* Nanoscale resolution in GFP-based microscopy. *Nature Methods* **3**, 721–723 (2006).

LETTERS

Pyrrolysyl-tRNA synthetase–tRNA^{Pyl} structure reveals the molecular basis of orthogonality

Kayo Nozawa[1]*, Patrick O'Donoghue[2]*, Sarath Gundllapalli[2]*, Yuhei Araiso[1], Ryuichiro Ishitani[4], Takuya Umehara[2], Dieter Söll[2,3] & Osamu Nureki[1,4]

Pyrrolysine (Pyl), the 22nd natural amino acid, is genetically encoded by UAG and inserted into proteins by the unique suppressor tRNA^{Pyl} (ref. 1). The Methanosarcinaceae produce Pyl and express Pyl-containing methyltransferases that allow growth on methylamines[2]. Homologous methyltransferases and the Pyl biosynthetic and coding machinery are also found in two bacterial species[1,3]. Pyl coding is maintained by pyrrolysyl-tRNA synthetase (PylRS), which catalyses the formation of Pyl-tRNA^{Pyl} (refs 4, 5). Pyl is not a recent addition to the genetic code. PylRS was already present in the last universal common ancestor[6]; it then persisted in organisms that utilize methylamines as energy sources. Recent protein engineering efforts added non-canonical amino acids to the genetic code[7,8]. This technology relies on the directed evolution of an 'orthogonal' tRNA synthetase–tRNA pair in which an engineered aminoacyl-tRNA synthetase (aaRS) specifically and exclusively acylates the orthogonal tRNA with a non-canonical amino acid. For Pyl the natural evolutionary process developed such a system some 3 billion years ago. When transformed into *Escherichia coli*, *Methanosarcina barkeri* PylRS and tRNA^{Pyl} function as an orthogonal pair *in vivo*[5,9]. Here we show that *Desulfitobacterium hafniense* PylRS–tRNA^{Pyl} is an orthogonal pair *in vitro* and *in vivo*, and present the crystal structure of this orthogonal pair. The ancient emergence of PylRS–tRNA^{Pyl} allowed the evolution of unique structural features in both the protein and the tRNA. These structural elements manifest an intricate, specialized aaRS–tRNA interaction surface that is highly distinct from those observed in any other known aaRS–tRNA complex; it is this general property that underlies the molecular basis of orthogonality.

Unlike the archaeal PylRS sequences, the bacterial versions are encoded in two separate genes. The *pylS* gene encodes the *D. hafniense* PylRS (DhPylRS) presented here, which includes a tRNA recognition domain (tRNA-binding domain 1), the conserved tRNA synthetase class II catalytic domain, the bulge domain and a carboxy-terminal tail, which is also involved in tRNA recognition (Fig. 1a). In Pyl-decoding bacteria, a second gene (*pylSn*) encodes a 110-residue polypeptide that is homologous (20% identity) to the amino-terminal domain of archaeal PylRSs (Supplementary Fig. 1). The hydrophobic nature of this domain reduces the solubility of PylRS[9] and encumbers crystallography[6,10]. Whereas both the DhPylRS and truncated versions of *M. barkeri* PylRS are active in aminoacylating tRNA^{Pyl} *in vitro*, only the full-length archaeal PylRSs, with a K_d about tenfold lower than that of the bacterial enzyme, had sufficient activity to support protein synthesis *in vivo*[9].

When *M. barkeri* PylRS and tRNA^{Pyl} were transformed into *Escherichia coli*, they functioned as an orthogonal pair in the heterologous environment[5,9]. We now show, both *in vitro* (Supplementary Fig. 2) and *in vivo* (Supplementary Fig. 3), that DhPylRS–tRNA^{Pyl} is an orthogonal pair with less enzymatic activity than its full-length archaeal counterpart, and that the addition of PylSn did not significantly enhance DhPylRS activity. Aminoacylation was performed with $N^{ε}$-cyclopentyloxycarbonyl-L-lysine (Cyc) because of the difficulty in synthesizing Pyl chemically[11]. DhPylRS acylates 80% of tRNA^{Pyl} transcript with Cyc, which is more than that of the homologous *Methanosarcina mazei* PylRS (MmPylRS) fragment (53%), and comparable to the 85% achieved by the full-length archaeal PylRS (Supplementary Fig. 2). Unfractionated *E. coli* tRNA is not a substrate for DhPylRS, and Cyc-tRNA^{Pyl} formation is not perturbed by competition with total *E. coli* tRNA. Similar results for *in vitro* aminoacylation by full-length and truncated MmPylRSs were reported recently[10]. In attempting to suppress a *lacZ* amber mutant, DhPylRS did not make enough Cyc-tRNA^{Pyl} to yield detectable β-galactosidase activity[9]. We therefore applied a strong selection[12,13] in which Cyc-tRNA^{Pyl} was required to suppress an *E. coli trpA* amber mutation and thus convert the test strain from Trp auxotrophy to prototrophic growth (Supplementary Fig. 3). We observed no growth on solid medium in the negative controls, and growth was observed only when PylRS, tRNA^{Pyl} and Cyc were present (Supplementary Fig. 3a–c). Although the addition of Trp to minimal liquid medium resulted in a wild-type growth rate (2.3 h doubling time), the *M. barkeri* PylRS–tRNA^{Pyl} (4 h) and the DhPylRS–tRNA^{Pyl} (5.8 h) had slower but significant growth with Cyc supplementation in the absence of Trp (Supplementary Fig. 3d).

To understand the molecular details of the PylRS–tRNA^{Pyl} interaction, we determined the crystal structures of the apo enzyme and of DhPylRS complexed to *D. hafniense* tRNA^{Pyl} at 2.5 and 3.1 Å resolution, respectively (Fig. 1 and Supplementary Tables 1 and 2). The protein forms a dimer in the crystal and in solution (data not shown). The final model of DhPylRS–tRNA^{Pyl} includes residues 10–288 of DhPylRS and tRNA^{Pyl}. The asymmetric unit of the complex crystal contains a DhPylRS dimer and two tRNA^{Pyl} molecules. Each tRNA^{Pyl} interacts predominantly with one subunit but also makes specific contacts with the other protomer (Fig. 1b). The tRNA-binding domain 1 and C-terminal tail are unique to PylRSs. The α1 helix of the tRNA-binding domain 1, the C-terminal tail and the bulge domain of the opposite subunit form a U-shaped concave structure that is complementary in shape to the acceptor helix and directs the 3′ terminus of tRNA^{Pyl} to the motif 2 loop (Arg 160–Asn 170) in the catalytic site (Fig. 1c and Supplementary Discussion). In addition to the core-binding surface (Fig. 1c), these unique protein structural elements contribute to the orthogonality of PylRS–tRNA^{Pyl}.

The aaRSs are found in two protein families that are distinguished by their evolutionarily unrelated catalytic core domains and by the way in which they bind opposing sides of the tRNA (reviewed in

[1]Department of Biological Information, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, B34 4259 Nagatsuta-cho, Midori-ku, Yokohama-shi, Kanagawa 226-8501, Japan. [2]Department of Molecular Biophysics and Biochemistry, [3]Department of Chemistry, Yale University, New Haven, Connecticut 06520-8114, USA. [4]Department of Basic Medical Sciences, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.
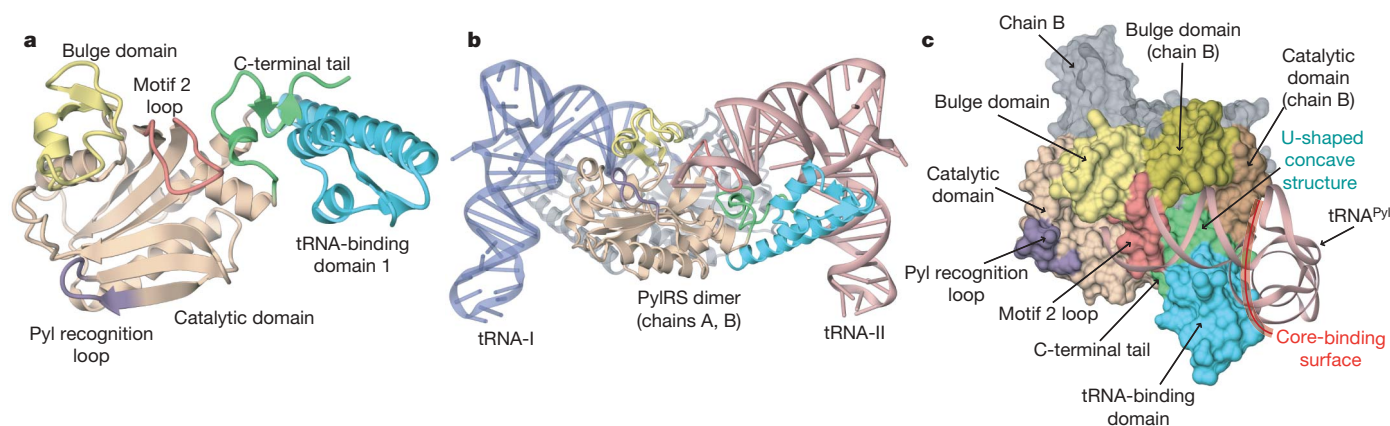*These authors contributed equally to this work.

Figure 1 | Overall structures. a, DhPylRS subunit A, shown as a ribbon model, consisting of the tRNA-binding domain 1 (1–68, blue), the catalytic domain (69–96 and 128–266, beige), the bulge domain (97–127, yellow), and the C-terminal tail (267–288, light green), with the motif 2 loop (160–170) coloured red and the Pyl recognition loop (212–218) coloured purple. **b**, The dimeric DhPylRS–tRNA[Pyl] complex structure, shown as a ribbon model. The asymmetric unit contains one PylRS dimer and two tRNA[Pyl] molecules: PylRS-A (coloured as in panel **a**), PylS-B (grey), tRNA-I (blue) and tRNA-II (pink). **c**, Binding of tRNA[Pyl] (ribbon representation) to the surface model of DhPylRS; structural domains are coloured as in **a**.

refs 14, 15). The class I aaRSs share a conserved Rossmann fold aminoacylation domain and (with the exception of TyrRS and TrpRS) approach from the minor-groove side of the tRNA acceptor stem. PylRS includes the conserved class II catalytic domain fold (only also observed in biotin synthetase and lipoyltransferase) and, like all other class II aaRSs, it approaches its tRNA from the major-groove side of the acceptor stem (Fig. 2). Similarities between the DhPylRS–tRNA[Pyl] and other class II aaRS–tRNA complexes do not extend much beyond these general features.

There are 31 protein residues in contact with the tRNA (annotated in Supplementary Figs 1 and 4). As in other aaRS–tRNA complexes that also lack the ATP or aminoacyl-adenylate substrate[16,17], the terminal adenosine (A76) occupies the ATP-binding pocket. In subsequent aaRS–tRNA complexes with ATP or aminoacyl-adenylate substrates, the terminal adenosine flips out of the ATP-binding pocket; this is accompanied by a slight conformational shift of the terminal CCA bases without affecting most of the interactions between the protein and tRNA[18,19]. Comparison with the MmPylRS–ATP complex[6] shows that only three residues (Arg 160, Leu 169 and Phe 172) are incompatible with simultaneous binding to ATP and the A76 adenylate.

Unique interactions between PylRS and tRNA[Pyl] contribute to orthogonality. Half of the 28 remaining residues, many of which participate in the core-binding surface, emerge from PylRS-specific

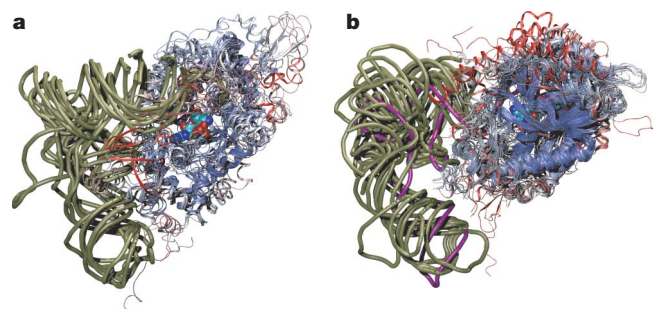**Figure 2 | Structural alignment of tRNA synthetase–tRNA complexes. a**, Class I; **b**, class II. Only a single monomer of the catalytic core domains is displayed, colour coded according to structural similarity. Viewed from the major groove side of the acceptor stem, a phosphate backbone outline of the tRNAs is shown (tan), and tRNA[Pyl] is shown in purple. In space-filling representation, a glutamyl-adenylate (**a**) and a pyrrolysyl-adenylate (**b**) highlight the class I and class II active site pockets, respectively. The small substrates are partly obscured by the protein backbone as a result of the need to show both aaRS families in the same orientation relative to the tRNA.
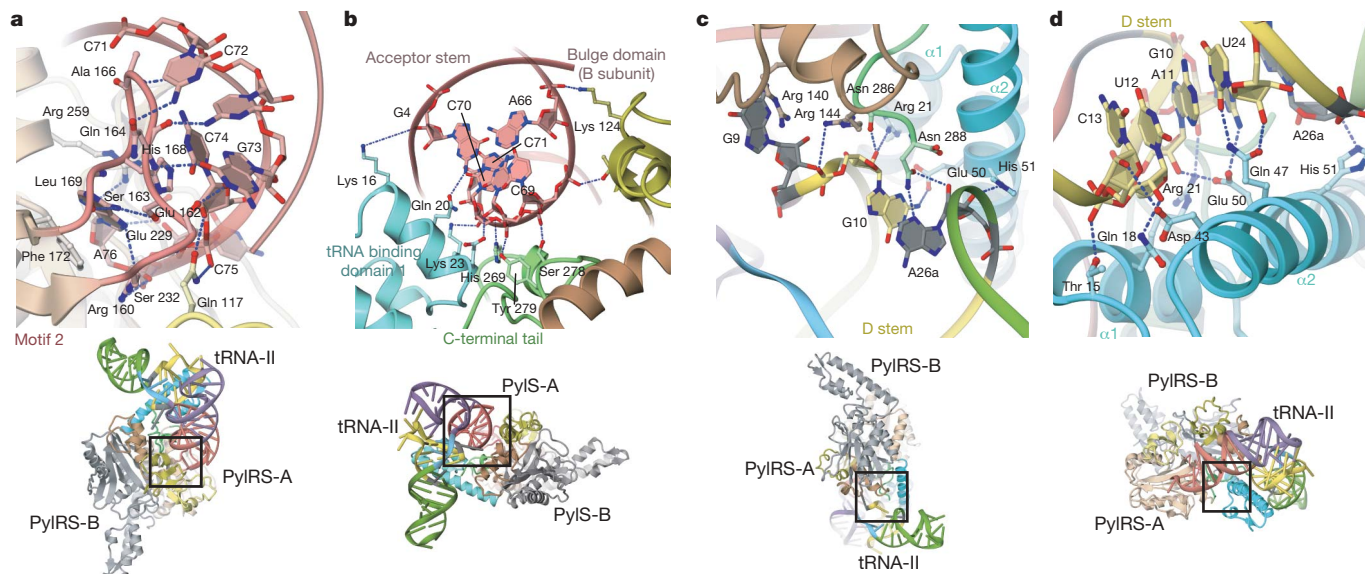
domains. Twelve residues from the class II catalytic domain and two residues from the bulge domain complete the tRNA-binding surface (Supplementary Fig. 4a). A comparison with the available class II aaRS co-crystal structures with completely docked tRNAs[18–23] indicates that PylRS uses typical tRNA-binding residues in unusual ways and also uses inserted residues—that is, those without homologous counterparts in other class II aaRSs—to make novel contacts with the tRNA. The imidazole ring of His 168, which is located in the motif 2 loop, participates in base stacking with C74 (Fig. 3a). In other class II aaRSs this position is occupied by His or Arg residues that invariably interact with C74 by means of hydrogen bonding. Gln 164 establishes three sequence-specific hydrogen bonds to C71, C72 and G73 (Fig. 3a). Only the yeast AspRS also places a tRNA-binding residue (Ser 329) at the homologous location, which establishes a single hydrogen bond to C74. Ser 163, an insertion in the motif 2 loop of PylRS, forms a hydrogen bond to the 6-amino group of A76 (Fig. 3a). Four contacts (Lys 124, Arg 140, Arg 144 and Glu 245) emerge from the B-chain of DhPylRS to contact the tRNA bound to the A-chain. Another inserted residue, Lys 124, interacts with the phosphate backbone at A66 (Fig. 3b). The side chains of Arg 140, Arg 144 and Glu 245 interact with G9 (Fig. 3c), which, peculiarly to tRNA[Pyl], is flipped outside the main tRNA body, so these interactions, too, are absent from other aaRS–tRNA complexes.

The tertiary core of a tRNA molecule is the location of sequence-distant interactions that are responsible for the canonical L-shaped tertiary structure of the tRNA. The core of tRNA[Pyl] is recognized by the core-binding surface, which is composed of tRNA-binding domain 1, the C-terminal tail and the α6 helix from the opposing protomer (Figs 1c and 3c, d). In tRNA[Pyl] the deletion of the otherwise invariant U8 base contributes to the compact core, as does an atypically short variable region and D-loop. Other unusual features of the tRNA[Pyl] sequence and secondary structure have been detailed elsewhere[1,24,25]. Deletion of U8 disrupts one of the most highly conserved tertiary base pairs (U8•A14) in tRNAs[26], leaving the non-standard guanosine at position 14 to base pair with C59 from the T-loop (Supplementary Fig. 5b, c). The U8 deletion also allows G9 to flip away from the tRNA body, where it is specifically recognized as a minor identity element[27] principally by means of a conserved cation–π interaction from Arg 140 (Fig. 3c). As a result of the absence of canonical position 48, a typical base pair between the T-loop and the D-loop is also missing (Supplementary Fig. 5b, c). Thus, these deletions in tRNA[Pyl] lead to a structurally rearranged and tightly packed tertiary core (Supplementary Discussion). PylRS evolved to form specific contacts with the compact core of tRNA[Pyl], including four strictly conserved contacts with the identity element base pairs

Figure 3 | DhPylRS–tRNA^Pyl interface. a, View showing the recognition of the CCA terminus by the motif 2 loop of DhPylRS. **b**, View showing the recognition of the tRNA^Pyl acceptor helix by DhPylRS. **c**, View showing the recognition of the tRNA^Pyl minimal core (acceptor–D junction region) by the core-binding surface of DhPylRS. **d**, View showing the recognition of the tRNA^Pyl minimal core (augmented D helix) by the core-binding surface.

G10•C25 and A11•U24 (ref. 27) (Fig. 3c, d). Interactions with the tertiary core, largely provided by tRNA-binding domain 1, make PylRS sterically incompatible with other canonical tRNAs because of their bulkier tertiary core.

Most of the interaction between PylRS and tRNA^Pyl is captured by the *D. hafniense pylS* gene product. The genetic code is maintained in all cellular life on Earth by the accurate aminoacylation of tRNAs with their cognate amino acids. In part, the fidelity of this interaction is due to tRNA identity elements; that is, those bases, base pairs and structural features of a tRNA that are crucial for aminoacylation by its cognate aaRS. The DhPylRS–tRNA^Pyl structure allows a more complete interpretation of tRNA^Pyl identity elements, elucidated by previous biochemical work. Identity elements for DhPylRS[27] include the discriminator base (G73), the first base pair in the acceptor stem (G1•C72), the D–stem base pairs G10•C25 and A11•U24, and G9, all of which are in direct contact with DhPylRS (Fig. 3a, c, d). Although

the full-length PylRS does not specifically recognize the anticodon, the two bases adjacent to the anticodon (U33 and A37) are identity elements for MmPylRS[25]. These two bases are possibly recognized in a sequence-specific manner by residues from the N-terminal domain (PylSn).

The tRNA-binding surface in DhPylRS shows a high degree of evolutionary conservation among PylRSs. Of the 28 tRNA-binding residues identified here, 15 (54%) are strictly conserved between all PylRS sequences, which compares with only 39% sequence identity between DhPylRS and its archaeal counterparts (Supplementary Figs 1 and 6). Five additional residues in the interface have generically conserved hydrogen-bonding potential, either through a similar side chain, for example Asp versus Asn at position 264, or because the hydrogen bond is through the protein backbone and is thus sequence independent, for example at Ala 166. Two residues (Lys 16 and Ser 278) are highly conserved, although not strictly so. The hydrogen
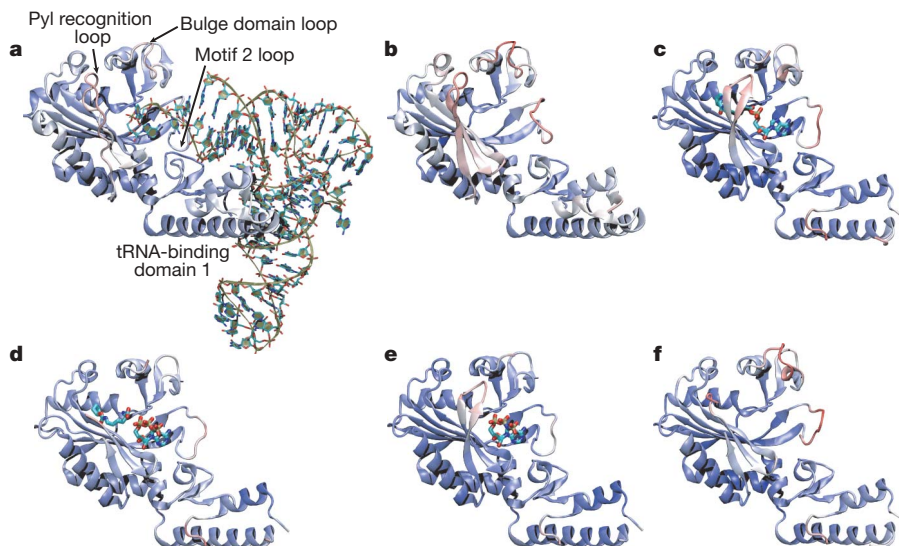

Figure 4 | Comparison of PylRS structures. a, DhPylRS–tRNA^Pyl complex (PDB code 2zni); **b**, DhPylRS apo (PDB code 2znj); **c**, MmPylRS–Pyl-AMP complex (PDB code 2zim)[6]; **d**, MmPylRS-ATP, Cyc complex (PDB code 2q7g)[6]; **e**, MmPylRS-ATP analogue complex (PDB code 2q7e)[6]; **f**, MmPylRS apo (PDB code 2e3c)[10]. The structures are coloured according to *B*-factor (indicating more (red) and less (blue) structurally dynamic regions), and four regions that show conformational changes in the different structures are labelled in **a**. Only one subunit of the dimer is shown for clarity.

bonds presented at these positions are probably somewhat less important interactions. Only four residues provide hydrogen bonds to the tRNA but are not conserved among PylRS sequences. Taken together, among PylRS sequences 81% of the residues in the DhPylRS–tRNA^{Pyl} interface are identical, are highly conserved or have conservation of an amino acid property that is important for the interaction such as charge or hydrogen-bonding potential. One residue (Glu 245), which is involved in a hydrogen bond network that contacts G9, is conserved only in the bacterial PylRSs. Archaeal tRNAs have a U at this position, and the smaller pyrimidine supports fewer contacts to the protein, which may contribute to the more robust aminoacylation by DhPylRS in comparison with a similar fragment of MmPylRS (Supplementary Fig. 2).

A recent attempt to identify the PylRS–tRNA^{Pyl} interaction by homology modelling suggested ten protein residues that might have a function in the interface[10]. Alanine scanning mutations showed that seven of these residues affected aminoacylation yields moderately to severely. The mutational analysis could not distinguish between mutants that alter protein stability and those that affect the stability of the interface between protein and nucleic acid. Of the 31 amino-acid residues in direct contact with tRNA^{Pyl}, only 6 were correctly predicted. The fact that homology modelling could not capture most of the PylRS–tRNA^{Pyl} interface highlights the distinctiveness of this complex in comparison with other aaRS–tRNA complexes.

In addition to the DhPylRS structures presented here, other structures are available for the homologous MmPylRS fragment[6,10]. Although MmPylRS and DhPylRS share only 39% sequence identity in this fragment of the molecule, the two proteins are highly structurally similar, displaying a root mean squared deviation of 2.4 Å and a structural similarity of 63% according to the metric Q_H (ref. 14). The only significant difference in the active site of the DhPylRS enzyme is the replacement of a bulky Trp 139 residue with the diminutive Leu 209 residue in MmPylRS, also recently observed in a DhPylRS apoenzyme structure[28]. Homology modelling shows that Pyl fits into the smaller active site of DhPylRS (Supplementary Fig. 7).

There are four regions with significant differences between the solved structures of PylRS (Fig. 4). First, tRNA-binding domain 1 is well ordered in the DhPylRS structures but not in the MmPylRS structures. This region seems to be more stable in general in the DhPylRS context, possibly explaining why the DhPylRS shows a higher aminoacylation yield than the homologous fragment of the *M. mazei* enzyme (Supplementary Fig. 2). Comparison of the DhPylRS structures shows that tRNA binding induces side-chain order as a result of specific interaction between the tRNA and tRNA-binding domain 1. Second, a conserved Tyr (DhPylRS Tyr 217; Supplementary Fig. 7) in the Pyl recognition loop (Fig. 1 and Supplementary Fig. 1) forms a hydrogen bond to the pyrrole ring nitrogen in the MmPylRS pyrrolysyl-adenylate complex[6]. In DhPylRS, as in other PylRS structures that lack the substrate Pyl, the Pyl recognition loop is not well ordered in our structures. The flexible nature of this loop was captured here, because one of the three asymmetric molecules in the C2 crystal was observed in the closed form whereas in the other molecules the loop was bent away from the empty active site. Third, on tRNA binding, the motif 2 loop alters its conformation drastically and becomes intercalated into the major groove of the acceptor end (Fig. 3a), providing base-specific recognition of the G1•C72 base pair as well as the discriminator G73 (Fig. 3a). Fourth, the bulge domain loop (residues 110–117) is resolved and well ordered only in the tRNA complex. Loop ordering is induced directly through a hydrogen bond between Gln 117 and the backbone phosphate oxygen of C75 and also indirectly as a result of interactions (for example, a hydrogen bond between Glu 162 and Gln 117) with motif 2, which orders on tRNA binding (Fig. 3a).

The class II aaRSs can be divided into three subclasses. Except for SerRS, the members of subclass IIa (HisRS, GlyRS α_2, ThrRS, ProRS and SerRS) and subclass IIb (AspRS, AsnRS and LysRS) include subclass-specific anticodon-binding domains. The homologous

domain architecture results in a mode of tRNA binding that is more similar among members of the subclass than between members of different subclasses. The amino acids genetically encoded by subclass IIb members are chemically similar (large polar or charged amino acids, namely Asn, Asp and Lys). The subclass IIa aaRSs are responsible for the small amino acids (Gly, Pro) and small polar amino acids (His, Thr, Ser). These trends are also observed among the subclasses of class I aaRSs[14]. Subclass IIc, including PylRS, PheRS, *O*-phosphoseryl-tRNA synthetase (SepRS), GlyRS $(\alpha\beta)_2$ and AlaRS, represent a different kind of evolutionary phenomenon. The amino-acid substrates are chemically more diverse than those in any other subclass, and they do not share homologous anticodon-binding domains. Comparison of the DhPylRS–tRNA^{Pyl} complex with tRNA complexes of other subclass IIc synthetases further shows that the detailed interactions between protein and tRNA are not conserved among subclass IIc members. The genesis of the class II aaRS family involved the initial radiation of three molecular lineages, namely the ancestral molecules from which the three subclasses evolved. Two of the three aaRS progenitors were restricted in their evolution, being selected to recognize large polar (subclass IIb) or small polar amino acid substrates (subclass IIa). The subclass IIc ancestor, from which PylRS evolved, was more adaptable than the other two, and perhaps its selective value is due to a greater inherent evolutionary plasticity.

METHODS SUMMARY

DhPylRS was overproduced and purified by published methods[9,27]. *D. hafniense* tRNA^{Pyl} was transcribed with T7 RNA polymerase, and purified under denaturing conditions by PAGE. For crystallization DhPylRS was mixed with tRNA^{Pyl} in a molar ratio of 2:2.2, at a final protein concentration of 5 mg ml^{-1}. The complex crystals grew at 20 °C by hanging-drop vapour diffusion against a reservoir solution of 90 mM MES-NaOH buffer pH 6.0 containing 5.4% propan-2-ol, 180 mM calcium acetate, 2% ethanol and 10 mM Tris-HCl pH 8.5.

Crystallographic, data-collection and refinement statistics are given in Supplementary Tables 1 and 2.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Srinivasan, G., James, C. M. & Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459–1462 (2002).
2. Krzycki, J. A. The direct genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.* **8**, 706–712 (2005).
3. Zhang, Y. & Gladyshev, V. N. High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm *Olavius algarvensis*. *Nucleic Acids Res.* **35**, 4952–4963 (2007).
4. Polycarpo, C. *et al.* An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc. Natl Acad. Sci. USA* **101**, 12450–12454 (2004).
5. Blight, S. K. *et al.* Direct charging of tRNA$_{CUA}$ with pyrrolysine *in vitro* and *in vivo*. *Nature* **431**, 333–335 (2004).
6. Kavran, J. M. *et al.* Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl Acad. Sci. USA* **104**, 11268–11273 (2007).
7. Wang, L., Xie, J. & Schultz, P. G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225–249 (2006).
8. Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding N$^\varepsilon$-acetyllysine in recombinant proteins. *Nature Chem. Biol.* **4**, 232–234 (2008).
9. Herring, S. *et al.* The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable *in vitro* but required for *in vivo* activity. *FEBS Lett.* **581**, 3197–3203 (2007).
10. Yanagisawa, T. *et al.* Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *J. Mol. Biol.* **378**, 634–652 (2008).
11. Polycarpo, C. R. *et al.* Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* **580**, 6695–6700 (2006).
12. Yanofsky, C. & Horn, V. Tryptophan synthetase chain positions affected by mutations near the ends of the genetic map of *trpA* of *Escherichia coli*. *J. Biol. Chem.* **247**, 4494–4498 (1972).
13. Murgola, E. J. tRNA, suppression, and the code. *Annu. Rev. Genet.* **19**, 57–80 (1985).
14. O'Donoghue, P. & Luthey-Schulten, Z. On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol. Mol. Biol. Rev.* **67**, 550–573 (2003).
15. Vasil'eva, I. A. & Moor, N. A. Interaction of aminoacyl-tRNA synthetases with tRNA: general principles and distinguishing characteristics of the high-molecular-weight substrate recognition. *Biochemistry (Mosc.)* **72**, 247–263 (2007).

16. Ruff, M. *et al.* Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA[Asp]. *Science* **252**, 1682–1689 (1991).

17. Goldgur, Y. *et al.* The crystal structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus* complexed with cognate tRNA[Phe]. *Structure* **5**, 59–68 (1997).

18. Cavarelli, J. *et al.* The active site of yeast aspartyl-tRNA synthetase: structural and functional aspects of the aminoacylation reaction. *EMBO J.* **13**, 327–337 (1994).

19. Moor, N., Kotik-Kogan, O., Tworowski, D., Sukhanova, M. & Safro, M. The crystal structure of the ternary complex of phenylalanyl-tRNA synthetase with tRNA[Phe] and a phenylalanyl-adenylate analogue reveals a conformational switch of the CCA end. *Biochemistry* **45**, 10572–10583 (2006).

20. Biou, V., Yaremchuk, A., Tukalo, M. & Cusack, S. The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with tRNA[Ser]. *Science* **263**, 1404–1410 (1994).

21. Eiler, S., Dock-Bregeon, A., Moulinier, L., Thierry, J. C. & Moras, D. Synthesis of aspartyl-tRNA[Asp] in *Escherichia coli*–a snapshot of the second step. *EMBO J.* **18**, 6532–6541 (1999).

22. Sankaranarayanan, R. *et al.* The structure of threonyl-tRNA synthetase-tRNA[Thr] complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* **97**, 371–381 (1999).

23. Briand, C. *et al.* An intermediate step in the recognition of tRNA[Asp] by aspartyl-tRNA synthetase. *J. Mol. Biol.* **299**, 1051–1060 (2000).

24. Théobald-Dietrich, A., Frugier, M., Giegé, R. & Rudinger-Thirion, J. Atypical archaeal tRNA pyrrolysine transcript behaves towards EF-Tu as a typical elongator tRNA. *Nucleic Acids Res.* **32**, 1091–1096 (2004).

25. Ambrogelly, A. *et al.* Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc. Natl Acad. Sci. USA* **104**, 3141–3146 (2007).

26. Marck, C. & Grosjean, H. tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* **8**, 1189–1232 (2002).

27. Herring, S., Ambrogelly, A., Polycarpo, C. R. & Söll, D. Recognition of pyrrolysine tRNA by the *Desulfitobacterium hafniense* pyrrolysyl-tRNA synthetase. *Nucleic Acids Res.* **35**, 1270–1278 (2007).

28. Lee, M. M. *et al.* Structure of *Desulfitobacterium hafniense* PylSc, a pyrrolysyl-tRNA synthetase. *Biochem. Biophys. Res. Commun.* **374**, 470–474 (2008).

METHODS

Additional crystallographic details. To crystallize DhPylRS alone, a 0.8-µl aliquot of the protein solution was mixed with 0.8 µl of a crystallization solution and then equilibrated by hanging-drop vapour diffusion against 500 µl of reservoir solution. Crystals appeared at 273 K within a few days in 45 mM sodium cacodylate buffer pH 6.0 containing 45 mM $(NH_4)_2SO_4$, 18% PEG400, 10 mM sodium acetate pH 4.6 and 0.2 mM sodium formate. To crystallize the DhPylRS–tRNA^{Pyl} complex, DhPylRS was mixed with tRNA^{Pyl} in a molar ratio of 2:2.2, respectively, at a final protein concentration of 5 mg ml^{-1}. The 0.8-µl aliquot of the complex solution was mixed with 0.8 µl of a crystallization solution and was equilibrated by hanging-drop vapour diffusion against a 500-µl reservoir solution of 90 mM MES-NaOH buffer pH 6.0 containing 5.4% propan-2-ol, 180 mM calcium acetate, 2% ethanol and 10 mM Tris-HCl pH 8.5.

For data collection under cryogenic conditions, the drop solution was slowly equilibrated against the same reservoir solution as for crystallization. The crystals obtained were briefly transferred to $1.2 \times$ reservoir solution containing 25% (w/v) PEG400 and were flash-cooled in a cryo-stream of nitrogen gas at 100 K. The data sets of the crystals were collected at station BL41XU of SPring-8 (Harima, Japan) and the NW12 beamline of PF-AR (Tsukuba, Japan). The data set of the DhPylRS–tRNA^{Pyl} was collected at 100 K in a cryo-stream of nitrogen gas. The collected data sets were processed with HKL2000 (HKL Research).

The complex crystal structure was solved by molecular replacement, using the free-form DhPylRS structure as a search model. Molecular replacement was done by MOLREP[29]; tRNA models were built with program O[30]. The solution was refined by CNS[31], by rigid-body refinement, energy minimization, B-factor refinement, and simulated annealing methods. In the resultant $2m|F_o| - D|F_c|$ and $m|F_o| - D|F_c|$ maps, we clearly found tRNA electron densities. Throughout the structural refinement, the WC base pair and the non-crystallographic symmetry restraints (weight = 300 kcal mol^{-1} Å^{-2}) were applied for the tRNA molecules. The crystallographic, data-collection and refinement statistics are presented in Supplementary Tables 1 and 2.

Aminoacylation assays. These assays were performed as described[32]. Aminoacylation reactions were performed at 37 °C in 100 mM Hepes-NaOH pH 7.2, 25 mM MgCl$_2$, 60 mM NaCl, 5 mM ATP, 1 mM dithiothreitol, 10 mM N^ε-cyclopentyloxycarbonyl-L-lysine (Cyc)[33] and 1 µM 3′-^{32}P-labelled tRNA. PylRS enzyme is added in excess at a concentration of 1 µM. In the aminoacylation assay including PylSn, DhPylRS and DhPylSn were added in an equimolar ratio (1 µM each). The Pyl analogue Cyc is used for aminoacylation assays. Reactions were stopped by removing 2 µl of reaction mix, adding it to 3 µl of 2.5 U ml^{-1} nuclease P1 (American Bioanalytical) in 0.1 M sodium citrate pH 4.5 at 25 °C for 30 min, thus releasing [α-^{32}P]AMP from uncharged tRNA, and Cyc-[α-^{32}P]AMP from charged tRNA. [α-^{32}P]AMP and Cyc-[α-^{32}P]AMP were separated by thin-layer chromatography on PEI cellulose plates (J. T. Baker) in 0.1 M sodium acetate and 5% acetic acid[34]. [α-^{32}P]AMP and Cyc-[α-^{32}P]AMP were exposed to PhosphorImager plates (FujiFilms), detected with a Molecular Dynamics Storm 860 scanner (Amersham) and quantified with ImagQuant software. Data are the results from at least three independent experiments.

Complementation of tryptophan auxotroph. To determine charging by DhPylRS *in vivo*, *E. coli* mutant strain FTP5822 (refs 35, 36) with an amber mutation in position 243 (Gln 243 to UAG) of TrpA was co-transformed with pCBS-PylS and pTECH-PylT. *PylS* was cloned into the NdeI and KpnI sites of pCBS (ampicillin marker), and *pylT* was cloned into the pTECH vector (chloramphenicol marker) as described[33]. In the complementation that included the *D. hafniense pylSn*, the *pylSn* gene was fused to the *lpp* promoter by PCR and then cloned into the pCBS-*pylS* vector between SmaI and BglII sites.

The experiment was performed as described previously[33]. PylS and PylT were under the control of the constitutive promoters *trpS* and *lpp*, respectively. The transformants were grown overnight at 37 °C in Luria–Bertani liquid medium. Subsequently, cells were washed thrice with M9 minimal medium and streaked onto M9 minimal agar plates supplemented with 19 amino acids (20 µg ml^{-1}) without tryptophan and 10 mM Cyc. Plates were incubated for 3–5 days at 37 °C.

Bioinformatic analysis. For comparison of the PylRS–tRNA^{Pyl} interface with other class II aaRS–tRNA complexes, protein structures were downloaded from the Protein Databank[34] and structurally aligned on the basis of the protein backbone of the conserved catalytic core domain by using Multiseq 2.0 in VMD 1.8.6 (ref. 37). The PDB codes for the aligned structures are as follows: class I aaRSs: 1f7v, 1gax, 1gtr 1n78, 1qf6, 1qu3, 1u0b, 1wz2, 2azx, 2ct8, 2cv2, 2dlc, 2dr2 and 2re8; class II aaRSs: 1asy, 1asz, 1c0a, 1efw, 1eiy, 1h4s, 1il1, 1il2, 1qf6, 1ser, 1wle, 1wyd, 1x54, 1yfs, 2cj9, 2du3, 2du4, 2du7, 2i40, 2iy5, 2j3l, 2odr and 2zim. Multiseq was also used for sequence alignment, alignment editing and calculation of sequence and structure conservation. Protein structures were downloaded from the National Center for Biotechnology Information and from the Joint Genome Institute Integrated Microbial Genomes with Microbiome Samples database[38].

Molecular graphics pictures in Figs 1 and 3 and Supplementary Figs 4 and 5 were prepared with the program CueMol (http://www.cuemol.org/). Figs 2 and 4 and Supplementary Figs 6 and 7 were prepared with the program VMD 1.8.6 (ref. 39).

29. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025 (1997).
30. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
31. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).
32. Herring, S., Ambrogelly, A., Polycarpo, C. R. & Söll, D. Recognition of pyrrolysine tRNA by the *Desulfitobacterium hafniense* pyrrolysyl-tRNA synthetase. *Nucleic Acids Res.* **35**, 1270–1278 (2007).
33. Polycarpo, C. R. *et al.* Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* **580**, 6695–6700 (2006).
34. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
35. Yanofsky, C. & Horn, V. Tryptophan synthetase chain positions affected by mutations near the ends of the genetic map of *trpA* of *Escherichia coli. J. Biol. Chem.* **247**, 4494–4498 (1972).
36. Murgola, E. J. tRNA, suppression, and the code. *Annu. Rev. Genet.* **19**, 57–80 (1985).
37. Roberts, E., Eargle, J., Wright, D. & Luthey-Schulten, Z. MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* **7**, 382 (2006).
38. Markowitz, V. M. *et al.* IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* **36**, D534–D538 (2008).
39. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

CORRIGENDUM

A role for VEGF as a negative regulator of pericyte function and vessel maturation

Joshua I. Greenberg, David J. Shields, Samuel G. Barillas,
Lisette M. Acevedo, Eric Murphy, Jianhua Huang, Lea Scheppke,
Christian Stockmann, Randall S. Johnson, Niren Angle
& David A. Cheresh

Nature **456**, 809–813 (2008)

In Fig. 2b of this Letter, the labelling of the *x* axis of the graph was incorrect. The corrected figure is printed below.
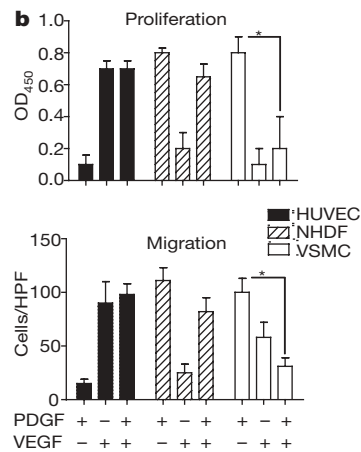
ERRATUM

Adaptive immune features of natural killer cells

Joseph C. Sun, Joshua N. Beilke & Lewis L. Lanier

Nature **457**, 557–561 (2009)

In the right panel of Fig. 4b of this Article the *x*-axis labels for 'Naive' and 'Memory' were inadvertently reversed in the histogram. The 'Naive' label should be on the left and 'Memory' on the right.

CORRIGENDUM

Regulation of *ERBB2* by oestrogen receptor–PAX2 determines response to tamoxifen

Antoni Hurtado, Kelly A. Holmes, Timothy R. Geistlinger, Iain R. Hutcheson, Robert I. Nicholson, Myles Brown, Jie Jiang, William J. Howat, Simak Ali & Jason S. Carroll

Nature **456**, 663–666 (2008)

In the Methods Summary, online-only Methods and the legend to Supplementary Fig. 16, the antibody catalogue number for the PAX2 antibody was incorrectly listed as ab38738. The correct antibody used for PAX2 immunohistochemistry was ab23799 (Abcam).

naturejobs

Last week *Nature* published a series of Commentaries discussing how science should react to the recession (see www.nature.com/recessionwatch). Prominent thinkers, policy-makers and scientists weighed in with incisive analyses about changes to science institutions, government programmes and economic tools that could help recovery. Within these forward-thinking treatises are potentially important tips for the individual scientist and his or her career.

Jeffrey Sachs, the director of the Earth Institute at Columbia University in New York, calls on the G20 countries to support sustainable energy in the poorest countries. This would be a "triple win" writes Sachs — a stimulus for richer countries, development for the poor countries, and environmental sustainability that would benefit everyone. Scientists should look to these initiatives as possible job generators.

Governments, such as that of Japan, should guard against becoming too insular with scientific funding and education, according to Atsushi Sunami and Kiyoshi Kurokawa, professors at the National Graduate Institute for Policy Studies in Tokyo. Nations, as well as individual scientists, need insights from abroad to battle mammoth issues, such as climate change and health.

For those in the corporate world, Noreena Hertz, a visiting professor of globalization at Erasmus University, the Netherlands, had a message: cooperate. Economic crises, such as the Great Depression, have fuelled cultural shifts towards social justice and communitarian values. During the current crisis companies should, therefore, become more proactive about advocating community projects devoted to the public good. Scientists working in industry or on start-ups should understand that public perception matters more than ever.

But regardless of scientists' career-track preferences, they must be willing to advocate for more investment in science and science education, according to UK member of parliament Ian Taylor. Too often, writes Taylor, scientists eschew advocacy and assume that the justification for support is self-evident. In this economic climate, that could be an egregious error.

Gene Russo is editor of *Naturejobs*.

Weathering the slump

Despite an economy that is sputtering, Ireland has recommitted to investing in science. **Quirin Schiermeier** examines the latest research investments on the Emerald Isle.

Scientists in Ireland held their breath last autumn as the financial storm that started in Manhattan hit their high-revving economy, buffeting the housing and construction sectors. It was a relief when, after a few tense weeks, the government redoubled its support for science.

Renewed investment in science and development play a key part in Ireland's recovery plan. "That the government has not blinked, despite the dire economic situation, sends a very strong signal to the Irish research community," says Frank Gannon, director of Science Foundation Ireland (SFI), a government-funded grant-giving agency. Its creation in 2000, with an influx of new funds, heralded the dawn of a new era for Irish science (see *Nature* **444**, 396–397; 2006).

In December, as an economic downturn seemed likely, the Taoiseach (prime minister) Brian Cowen reaffirmed a previous commitment to invest some €8.5 billion (US$10.6 billion) in science over the next six years. In its framework for sustainable economic revival, the government outlined 45 key actions — from a €500-million venture fund to support research and development by early-stage companies, to fast-track visas for foreign researchers and their spouses — aimed at turning Ireland, with its population of 4.2 million, into an 'island of innovation'.

With cash in hand, and newly introduced peer-review standards to help ensure quality research, the SFI helped to start a remarkable upswing that attracted talent from abroad and kept Irish scientists in the country. Funding is available in three government-established growth areas considered vital for Ireland's future: the life sciences and biotechnology, information and communication technologies, and sustainable energy and energy-efficient technologies. Its annual €200-million budget will continue to grow at the same rate, around 4%, as in previous years. This, says Gannon, should be sufficient to keep up the healthy 25% success rate of applicants for individual SFI grants, including 'research frontiers' grants for more risky science, 'starting investigator' grants for early career scientists, and principal-investigator grants worth €250,000 per year on average.

The SFI's funding activities, through government agencies such as Enterprise Ireland and Industrial Development Agency Ireland, are part of a broader

Fergus Shanahan and Brian MacCraith see the benefits of industry partnership.

strategy aimed at transforming Ireland's manufacturing sector. Some 40% of SFI-funded investigators have active connections with industry. Overall, 90% of Ireland's public-science funding is earmarked for research with potential commercial applications in an attempt to maximize jobs and economic growth.

Despite these programmes, Ireland's research expenditure is still relatively modest. At around 1.4% of gross domestic product, research spending in the republic is considerably below the European Union (EU) average, and even slightly below that of China. But in terms of numbers of researchers, Ireland, with 5.7 researchers per thousand in the labour force, is now ranked slightly above the EU average (5.6), and is almost on par with Switzerland (5.8) and Britain (6.2), according to 2006 figures just released by the EU.

Small size has its virtues. "We're able to have a census of all groups we support," says Gannon. "That's a real advantage when it comes to coordinating our research efforts." All researchers applying for SFI grants must demonstrate how the project could have an economic impact. Applications for the flagship scheme, the SFI Centres for Science, Engineering and Technology (CSET), through which large academic–industrial partnerships can receive up to €5 million per year, undergo particularly thorough review. Centres are expected to focus on research with commercial potential and to deliver publications and patents.

Among the CSET beneficiaries is Fergus Shanahan, director of the Alimentary Pharmabiotic Centre at University College Cork, which is run in partnership with GlaxoSmithKline. He recently won a second round of funding. Over the next five years, the SFI will support the centre — one of nine ventures funded through the CSET programme — with an overall €17.5 million; an additional €5 million will come from industry. "It's a blessing in these times that the government has reconfirmed so determinedly its commitment to science," says Shanahan. "I do hope we can repay the trust over the next five years." He expects opportunities for geneticists, microbiologists, clinicians and others.

All 56 group leaders and young investigators who have moved through the centre since its creation five years ago are still in science, most of them in Ireland, and most of them have moved up the career ladder. Around one-half are now working for industry. The

"If you have the right company and the right idea there is still plenty of seed money available."
— Kingston Mills

centre itself, says Ferguson, has become more international. The number of non-Irish natives among its 125 staff has doubled to 30% as more scientists from Europe, North America and Asia-Pacific have been recruited, suggesting that Ireland continues to be a draw. "Salary has long ceased to be a recruitment hindrance," says Shanahan.

The hub of activity

Dublin hosts much of the action. Last year, Ireland's first official nanoscience institute, the SFI-funded Centre for Research on Adaptive Nanostructures and Nanodevices, moved its 100 or so researchers into a stylish five-story building complete with an art gallery. The centre, is, among other things, developing nanoscale biosensors for virus detection.

Three new CSET centres opened their doors in Dublin last year. One was the Centre for Next Generation Localisation, led by Josef van Genabith of Dublin City University. Working with IBM and Microsoft, the centre is adapting computer software to different languages and regional cultures. And CLARITY, led by Barry Smith of University College Dublin, aims to improve sensor technology for wireless transmission of messages. Such sensors could be used to transmit large environmental data sets. Industrial partners include Ericsson, Vodafone and IBM.

Setting up a CSET centre, which typically employs 100–150 scientists, requires at least 25% involvement from industry in the form of money, personnel or both. These partnerships have various arrangements. At the

Size is important: Ireland's size allows Frank Gannon to better coordinate research.

Biomedical Diagnostics Institute at Dublin City University, for example, researchers from participating companies, including the biotech company Amic from Uppsala, Sweden, and medical technology company Becton Dickinson, headquartered in Franklin Lakes, New Jersey, are working with colleagues from academia. "This creates a very productive and innovative work environment," says Brian MacCraith, an optical physicist who runs the centre, which was established in 2005 with a €16.5-million SFI grant.

Arranging industry participation can take time, because it requires agreements on funding, patents and royalties. So in 2007, the SFI created the Strategic Research Clusters programme to back commercially promising ideas, giving investigators more time to attract and cultivate industry partnerships. Twelve such clusters have been announced so far, each funded by up to €1.5 million per year over a five-year period, and focusing on topics ranging from solar energy conversion to dairy-cow fertility.

Kingston Mills, an experimental immunologist, leads one such cluster, the Immunology Research Centre at Trinity College Dublin. It is a €10-million, five-year research programme. Since funding started early last year, he has recruited one group leader, nine postdocs and five PhD students whose research focuses on the discovery and function of novel activators and inhibitors of innate immunity. Private partnerships with Opsona Therapeutics, a Dublin-based spin-off company for vaccine development that Mills started in 2004, and the pharmaceutical firm Schering-Plough contribute some 20% to the overall funding. "Raising funds has become more difficult," he says. "But if you have the right company and the right idea there is still plenty of seed money available."

Meanwhile, Enterprise Ireland has announced that it will invest €20 million in a new research centre for functional foods designed to provide health benefits beyond basic nutrition. And a €150-million Biosciences Institute is being built in a public–private partnership involving Trinity College Dublin. Its planned opening in 2011 should provide collaboration opportunities for PhD students.

Ireland's government is seeking to double the number of PhD students over the next five years. But a decline in core funding — money that the ministry of education gives to Ireland's seven research universities — could affect this, says Mills. These funds cover running costs, such as electricity, building repairs and salaries. Fixed university budgets are to be cut further, by approximately 7% this year. "Grant money keeps increasing while core funding is eroding," says Mills. "There's a disconnection."

To bridge the gap, some government officials and university rectors favour the re-introduction of undergraduate student fees, which were abolished more than a decade ago, to provide greater access to higher education. Irish students hate the idea, but its champions argue that free education is no longer justified when so many Irish families can now easily afford to pay for it.

More families, however, could feel the pinch as the economy continues to cool. But the Irish economy has acquired a serious affinity for science — an affinity that should provide ample research opportunities for students and scientists for the foreseeable future. ■

Quirin Schiermeier is *Nature*'s Germany correspondent.

MOVERS

John Polich, director, Global NeuroLab Operations, NeuroFocus, Berkeley, California

1996–2009: Associate tenured professor, molecular and integrative neurosciences department, Scripps Research Institute, La Jolla, California
1994–2009: Adjunct professor, psychology department, University of California, San Diego, La Jolla, California

John Polich learned to drive a tractor on his family's Iowa farm, but he had little interest in farming. Looking at the field's lone cottonwood tree, he was more intrigued with how his brain deciphered that the object was a tree, or that it was green. Polich spent the next 30 years in academia as a cognitive neuroscientist studying memory and the brain's response to disease and addiction. Now, with a late career move into industry at NeuroFocus, he hopes to make use of the brain-wave activity techniques he's helped to develop.

Polich's propensity for contemplation led him to a Benedictine seminary in Conception, Missouri, for college. But after two years, he decided he wasn't monk material. He transferred to the University of Iowa, where he majored in experimental psychology. "I wanted to understand how the brain produced cognition," he says. In just eight months, he completed a master's degree in cognitive psychology at Wayne State University, Michigan.

As one of the first PhD candidates in cognitive psychology at Dartmouth College, New Hampshire, Polich did some of the earliest recording of brain waves. He later did a postdoc at the University of Illinois with Emanuel Donchin, a leader in the use of specific brain-wave signals to assess attention and memory.

Keen to deal with patients, he built his own lab at the Scripps Research Institute in La Jolla to study patients in the early stages of Alzheimer's disease. But, as he was not a physician, he found it difficult to recruit large numbers. So he changed his focus and spent 20 years studying how drugs such as marijuana, tobacco, caffeine and ecstasy affect neurotransmitters.

The increasing struggle for grant money helped lure him to NeuroFocus, a company interested in applying his brain-wave techniques to marketing-effectiveness studies — otherwise known as 'neuromarketing'. "Finding a whole new way to apply brain-wave techniques is an exciting and challenging career move," he says. Indeed, NeuroFocus is expanding to meet demand for neuromarketing expertise.

Bob Knight, a neuroscientist at the University of California, Berkeley, and chief scientific adviser to NeuroFocus, says Polich's work in attention, emotion and memory covers three cardinal facets of neuromarketing. "John is regarded as a meticulous researcher, and as we use science to help grow this industry, he and his impeccable scientific record will help us make sure our research is conducted at the highest level possible," says Knight. ■
Virginia Gewin

NETWORKS & SUPPORT

EMBL offers lab training in Europe

Cutting-edge lab training courses are coming to Europe. The European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany, is building a 16,000-square-metre Advanced Training Centre with a 450-person conference room, teaching labs, seminar rooms and an exhibition area.

Due to be completed in September, it will offer conferences and workshops on topics such as identifying protein structures, using biological databases and learning new imaging techniques. Although EMBL emphasizes graduate and postdoc training, the centre will hold conferences and workshops for scientists at all career stages.

EMBL sees this as the European counterpart to Cold Spring Harbor Laboratory (CSHL) in New York, says Matthias Hentze, associate director and coordinator of the centre. "Europe lacks such a centre," he adds. EMBL will supplement, not supplant, the CSHL's renowned programmes, he notes. Many courses and conferences at the CSHL are over-subscribed, or can be offered only intermittently. Some conferences will be available at both facilities, allowing more participants in hot fields. EMBL's central location will also save European scientists travel time and costs, Hentze says.

EMBL has also found a way to help visitors save some money. Leica Microsystems of Wetzlar, Germany, has committed €50,000 (US$63,000) per year, and Perkin-Elmer of Waltham, Massachusetts, €25,000 a year, for courses and conferences at the centre for its first three years. This is important because of the research and development funding disparity in the European Union, says Hentze. "There are many places in Europe where funding is very, very small — particularly in eastern Europe and some countries in the south. We will be able to host conferences at a low fee," he says.

Hentze says that support from the two firms and other anticipated sponsors will keep conference fees to €400–500 per person and pay entire costs for 200 fellows. "To maximize impact, we will probably use this support for partial and complete 'attendance fellowships' for the needy, rather than spreading it evenly," he says.

The training labs will use equipment from both Leica and Perkin-Elmer. Jörg Fleckenstein, senior manager of resource development at EMBL, says that "they can showpiece their equipment in their training courses, if scientific rationale supports it". He expects that additional sponsors will provide funds for more fellowships. ■
Paul Smaglik

POSTDOC JOURNAL

Nice day for a white wedding?

So much for my plan. I had every intention of facing the challenges of the new year with serene Zen-like calm, but I have already crumbled like a stale biscuit. In my defence, the challenges have been stacking up rapidly of late.

We returned home from a New Year camping trip to find my partner, Brett, had been offered a two-year postdoc in Colorado. After the euphoria, we quickly realized that we have only a few months to pack up our life, uproot our three-year-old son, and organize visas and flights for our overseas move.

However, our problems entered a whole new dimension when we discovered that our relationship is not recognized by the United States for visa purposes. For me to accompany Brett to the United States we have to get married — now!

So, in addition to working three separate postdoc contracts, which requires exceptional multitasking skills and many more hours than there are in the week, I now have to organize my nuptials. Although I am overjoyed for Brett, and excited about moving, I can't help but worry about my own career. As I plan to get married and support Brett in his new position, will my career ebb away while his takes off?

Perhaps I should take my own advice, and view this challenge as an opportunity to grow. Maybe I should even think seriously about those other, more child-friendly careers I've heard so much about. ■
Joanne Isaac is a postdoc in climate-change effects on biodiversity at James Cook University in Townsville, Australia.

Lost in sun and silence

The golden age of communication.

Vincenzo Palermo

Finally, he was alone.

The Sahara was black, covered by billions of dark solar panels, silently chasing the sun. He had to bribe the Libyan maintenance workers to leave him at the exact centre of African Phoenix, the largest fully robotic solar plant in the world.

Alone.

Well, not exactly.

"*I really think this is insane,*" said Rick's voice into his head.

"*Everybody thinks this is insane,*" commented Mei Li, another voice in his ears, "*but you know how Liam jumps into things. Let him stay there for an hour or two, and he will return home crying.*"

Looking at the display of his phone, Liam saw that most of his friends were online. They were there for him, to see if he really was going to do it. Rick was connected from his office in New York, Mei Li was chatting from home in Hong Kong; then Michael, Alan, Nikolaj … all his chat friends scattered around the world.

"*If you want to stay disconnected for some time, you just turn off your computer and phone, you don't go to the middle of nowhere,*" said Rick.

"How many different windows do you have open on your computer now, Rick?" asked Liam.

"*What do you mean?*"

"How many other things are you doing while talking with me, my friend?"

"*Come on, this is not important.*"

But that was exactly it. Liam knew he had the attention of all his friends at that moment, but none was giving him a 100% of their time. Liam was hardly in a position to criticize. His house in London was the first to have a TV in the bathroom, to keep tedium at bay, even in those special moments.

"I can't remember myself doing only one thing at a time. Ever," said Liam. "I can't remember working without checking my e-mail, or jogging without listening to music. Come on, Mei, even when we had sex, we had music on and our mobiles ringing!"

Mei Li didn't answer. Maybe she was chatting with somebody else.

That was exactly the reason why Liam, that same morning, took a low-cost flight from London to Timbuktu. The crowd of the city, where half of the solar energy feeding the world was produced, was the last bit of hyper-connection that Liam could stand. Turning off the phone is not enough (he said to everybody); it's like swimming in shallow water, where you can put your feet on the ground whenever you like. He didn't want any emergency escape.

"*Well,*" said Rick, "*call me when you are over with this bull …*"

He turned off his smartphone, his only remaining connection to the world. Then, he threw the phone into the air as hard as he could. It described a long arc and hit a silicon panel, bounced a couple of times and fell down in the underworld, between the solar cells and the sand.

Then, Liam discovered the Silence.

Nobody around for kilometres; no Internet connection; no way to communicate. He could hear the wind passing through the panels, whistling around the tubes, noises he hadn't noticed before. The huge amount of electricity produced and carried around seemed to induce a low hum in the air, or maybe it was just the noise of the piezoelectric actuators continuously aligning the panels to the sun.

Alone, and in silence — at last.

After some time, Liam got bored. There was nothing to do, just walk around and get sunburn. The desert was fascinating, but only for the first five minutes. Now he had shown he could do it, he could go home, talk about it at parties, and girls would ask: "Really? You stayed disconnected and alone for a *whole day*?"

The biggest fear of the twenty-first century was not death, but loneliness.

And low connection speed.

After some more time, he was not just bored, but thirsty and tired. The Libyan workers should come and take him back at 8 p.m., but without his mobile he had no watch, and couldn't tell the time. This drove him crazy. Mei Li was right. He was going to come back home and cry.

After what seemed a huge amount of time, he was in full crisis. What if the workers didn't come and rescue him? The sun was still high, but he was sure that 8 o'clock had come and gone. He could have called for help — before the stupid idea of throwing away the phone.

He ventured beneath the solar panels, away from the main road. The panels were mounted on poles at least two metres high, so he could walk under them in a pleasant shadow. It was easy to find the phone, lying in the soft sand, undamaged by the hard landing.

He had just turned on the phone when they came.

Four young guys, dressed hip-hop style, their clothes whitened and worn by a long stay in the desert. They all had long metallic tubes, pointed as spears. Very primitive. And scary. A gang of teenage marauders, as there were in every big city, but how the hell could they survive in the desert?

He tried to say: "Hello!"

They got closer. Their noses were deformed by cotton plugs soaked in gasoline, and they had headphones shooting music at maximum volume, keeping them in permanent hallucination. Liam turned back and ran away. He hadn't gone ten metres before the spear hit him. Falling down, he could hear Rick's happy voice, answering his call.

"*I had bet ten bucks on you calling back,*" said Rick. "*You could not survive a whole day alone.*"

■

Vincenzo Palermo is a chemist and a writer at the National Research Center, Bologna, Italy. He has published 40 scientific papers, mostly on nanotechnology and photovoltaics, several short stories and a historical novel on the Second World War.